Epigenetics in cellular reprogramming and cancer - investigation of DNA methylation barriers

Manasterski, Piotr

Award date:
2021

Awarding institution:
University of Bath

Epigenetics in cellular reprogramming and cancer - investigation of DNA methylation barriers

Piotr Jan Manasterski

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

November 2020

“He who has a why to live can bear almost any how”

- Friedrich Nietzsche

Copyright notice

Declaration of authorship

I am the author of this thesis, and the work described therein was carried out by myself personally, with the exception of the following:

Section 2.2.1 The passaging and maintenance of the H9 hESC cell line was performed by Dr Stephen Weston and Dr Zoë Burke from the Department of Biology and Biochemistry at the University of Bath

Section 2.4.1 The differentiation of the H9 hESC cell line to HLCs was performed by Dr Stephen Weston from the Department of Biology and Biochemistry at the University of Bath

Section 2.4.4 The FACS analysis was performed with the assistance of Dr Michael Zachariadis from the Material and Chemical Characterisation Facility at the University of Bath

Section 2.5.4 The work involving the MaXis HD ESI-QTOF mass spectrometer was performed with the assistance of Dr Shaun Reeksting from the Material and Chemical Characterisation Facility at the University of Bath

Section 2.5.5 The work involving the Xevo TqD triple quadrupole mass spectrometer was performed with the assistance from Dr Jack Rice and Andrew Kannan, under the supervision of Professor Barbara Kasprzyk-Hordern from the Department of Chemistry at the University of Bath

Section 2.5.5 The external method validation experiment was conducted by Dr Anna Caldwell at King's College London

Section 2.6.3 The assessment of *TET* transcript levels in a panel of CRC cell lines was performed by Tracey Collard at the University of Bristol

Section 2.8.3, section 2.8.4 and section 2.8.5 The experiments were conducted and analysed by Dr Jennifer Pinnell from the Department of Biology and Biochemistry at the University of Bath

Acknowledgements

First and foremost, I would like to thank my primary supervisor Professor Adele Murrell for the opportunity to study a fascinating topic and further develop my passion for cancer research. Thank you for all your help, advice and support. I will miss your dark humour (your jokes are safe with me) and Friday morning breakfasts. I would like to thank my secondary supervisor Professor David Tosh for giving me a chance to study stem cell differentiation, his encouragement and constructive feedback. I am also very grateful you were too preoccupied with running the Department to scare me more often when I was working in the lab.

I would like to thank Professor Barbara Kasprzyk-Hordern, Dr Jack Rice, Andrew Kannan and Dr Shaun Reeksting for their help with developing and optimising the mass spectrometry method as well as running the samples. I want to thank Professor Ann Williams and Tracey Collard for their assistance in selecting the colorectal cancer cell lines I used in my work. A big thank you to Dr Giuseppina Pisignano for putting up with me all these years, her help and good humour. I hope you don't miss me too much. Thanks to the members of the Murrell, Vance and Tosh lab groups: Dr Jennifer Pinnell, Dr Zoë Burke, Dr Stephen Weston, Dr Elizabeth Coe, Dr Ioanna Pavlaki and Stephen Richer for their contributions to my experimental work.

I would like to thank Chris Vennard for an amazing four years together and his invaluable hair grooming tips. Big thank you to Sandra Catalan for her continued encouragement, kind words and the wine jam from Spain. It was delicious.

A massive thank you to my training partner, Carmen Grimaldos, for all the laughs and crazy training sessions. I will miss you tremendously and hope to see you again on many triathlon start lines.

I would like to wholeheartedly thank Rugile Siaudinyte for her love and support during writing of this thesis. I could not have done it without you by my side.

The most important thanks go to my parents. Your unconditional love, warmth, encouragement and unyielding belief in me continue to inspire me to better myself every day. I am immensely grateful for that and never take it for granted. We can finally fulfil my life-long dream of all addressing each other as "Doctor" every time we meet.

Abstract

The discovery of ten-eleven translocation (TET) enzymes-mediated 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) gave rise to a new area of epigenetic study. Initially thought to be simply an intermediate in an active DNA demethylation pathway, 5hmC was since demonstrated to play critical roles in mammalian development, cellular differentiation and cancer progression. The global 5hmC content is reduced in the vast majority of studied cancers and mutations in the three *TET* genes are present to a varying extent in some tumour types. The reactivation of the TET proteins was proposed as a therapeutic intervention in melanoma patients. However, the preliminary data from the Murrell lab suggests that 5hmC levels increase during colorectal cancer (CRC) metastasis to the liver and few *TET* mutations are reported in CRC patients. Given that 90% of cancer deaths occur due to metastasis, the role of TETs and 5hmC in CRC progression warrants closer investigation.

The in-house ultrasensitive liquid chromatography-mass spectrometry method of global cytosine, 5mC and 5hmC levels detection in various DNA samples was successfully optimised and validated. It was found to robustly detect the 5hmC content in almost all DNA samples studied and offer a superior sensitivity to an external method. The primary SW480 and metastatic SW620 CRC cell lines were chosen to study the effects of 5hmC and TET levels manipulation. The SW480 cells were shown to have higher migration rate and *TET2*, 5mC and 5mhC content but lower *TET1* expression and proliferation rate compared to the SW620 cells. The treatment of the two cell lines with vitamin C increased their 5hmC content, but its phenotypic effects were found to be 5hmC-independent. The CRISPR/Cas9-based targeting of the three *TET* genes in these cell lines caused an incomplete ablation of TET and 5hmC levels and resulted in lower proliferation rate in the SW480 cells and reduced migration in the SW620 cells. The changes in *TET* expression and 5hmC levels are also described during the differentiation of human embryonic stem cells (hESCs) to hepatocyte-like cells (HLCs). Despite the initial attempt to knockdown *TET1* levels in hESC to improve differentiation efficiency failing, future experiments utilising this approach could produce HLCs closer resembling human hepatocytes, which is critical for drug testing purposes. Taken together, these results contribute to the growing body of evidence for the critical role of 5hmC and TETs in cancer progression and hESC differentiation.

Table of contents

Table of figures

Table of tables

List of abbreviations

5caC	5-carboxymethylcytosie
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AA	Ascorbic acid
AID	Activation-induced deaminase
AIML	Angioimmunoblastic T-cell lymphoma
AML	Acute myeloid leukaemia
APC	Adenomatous polyposis coli
APCI	Atmospheric pressure chemical ionisation
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide
BER	Base excision repair
C8	Octyl carbon chain
C18	Octadecyl carbon chain
CIMP	CpG island methylator phenotype
CGI	CpG island
CpG	Cytosine followed by a guanine on the same DNA strand
CRC	Colorectal cancer
CRISPR/Cas9	Clustered regularly interspaced short palindromic repeats/CRISPR-associated 9
CSC	Cancer stem cell
DE	Definitive endoderm
DMαKG	Dimethyl alpha-ketoglutarate

DMR	Differentially methylated region
DNMT	DNA methyltransferase
DKO	Double knockout
ED	Embryonic day
ESI	Electrospray ionisation
FACS	Fluorescence-activated cell sorting
FAP	Familial Adenomatous Polyposis
HB	Hepatoblast
HDAC	Histone deacetylase
HDM	Histone demethylase
hESC	Human embryonic stem cell
HILIC	Hydrophilic interaction liquid chromatography
HLC	Hepatocyte-like cell
HMT	Histone methyltransferase
HPLC	High-performance liquid chromatography
ICR	Imprinting control regions
IDH	Isocitrate dehydrogenase
iPSC	Induced pluripotent stem cell
ISC	Intestinal stem cell
KAP	KRAB-associated protein
KRAB	Krüppel-associated box
LC-MS	Liquid chromatography coupled to mass spectrometry
MALDI	Matrix-assisted laser desorption/ionization
MDS	Myelodysplastic syndrome

MeCP	Methyl-binding protein
mESC	Mouse embryonic stem cell
miRNA	Micro RNA
MRM	Multiple reaction monitoring
ncRNA	Non-coding RNA
NiCl2	Nickel (II) chloride
O-GlcNAc	O-linked N-Acetylglucosamine
OGT	O-GlcNAc transferase
PARP	PolyADP-ribose polymerase
PGC	Primordial germ cell
QC	Quality control standard
SKO	Single knockout
TA	Transit-amplifying
TDG	Thymine DNA glycosylase
TET	Ten-eleven translocation
TKO	Triple knockout
TOF	Time-of-flight
TSS	Transcription start site
UHRF	Ubiquitin-like with PHD and ring finger domains
ZFP	Zinc-finger protein

Chapter 1

General introduction

1.1 The times they are a-changin' - going beyond the DNA sequence through the rise of epigenetics

The idea of epigenetics has first been proposed by Conrad Waddington in his landmark 1942 paper as a way to describe the mechanistic relationship between the phenotype and the genotype of an organism. The term "epigenetics" in that setting described a set of causal mechanisms linking the actions of genes to the phenotype (Waddington, 1942). From Waddington's perspective, the study of epigenetics was mainly to be conducted with relation to the developmental processes and informed by the findings from the field of embryology. Indeed, this line of thinking has led him to create the idea of an "epigenetic landscape", portrayed in Figure 1.1. This concept refers to a series of choices that a cell has to make on its way towards a terminally differentiated state and in the process acquiring characteristics specific to that differentiated cell type. It allows for easy conceptual visualisation of the epigenetic barriers that accumulate during the differentiation process and prevent a cell, under normal circumstances, from changing its identity by dedifferentiating or transdifferentiating.

More recently, the concept of epigenetics has been transformed to refer to heritable changes in gene expression that are not dependent on alterations to the underlying DNA sequence. In other words, epigenetics is what allows cells with the same genotype to exhibit profound phenotypic differences. Since the original description of epigenetics, numerous components of the epigenetic machinery have been described. These include direct modifications of DNA bases such as N^6-methyladenosine, 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), 5-carboxycytosine (5caC), 5-hydroxymethyluracil and 5-formyluracil (reviewed in detail by Bilyard et al., 2020). The N-terminal tails of the histone proteins, around which the DNA is wrapped to form the nucleosome, can be modified by adding numerous chemical groups including phosphorylation, acetylation, methylation, sumoylation and ubiquitylation (reviewed by Bannister and Kouzarides, 2011). The addition and removal of these chemical groups in turn both change the structure of the surrounding chromatin (formed of the nucleosome subunits) and affect binding of proteins that contain domains responsible for recognising these modifications. Lastly, non-coding RNAs (ncRNAs) form the third main component of the epigenetic machinery (reviewed by Mongelli et al., 2019; Yang et al., 2020). The ncRNAs are transcribed but not translated to form a

protein. They hence act as an independent class of signalling molecules involved in numerous cellular processes such as regulating chromosomal architecture (reviewed by Creamer and Lawrence, 2017) and alternative splicing of RNAs (reviewed in detail by Romero-Barrios et al., 2018). They include, but are not limited to, circular RNAs, long ncRNAs, micro RNAs (miRNAs) and small interfering RNAs.

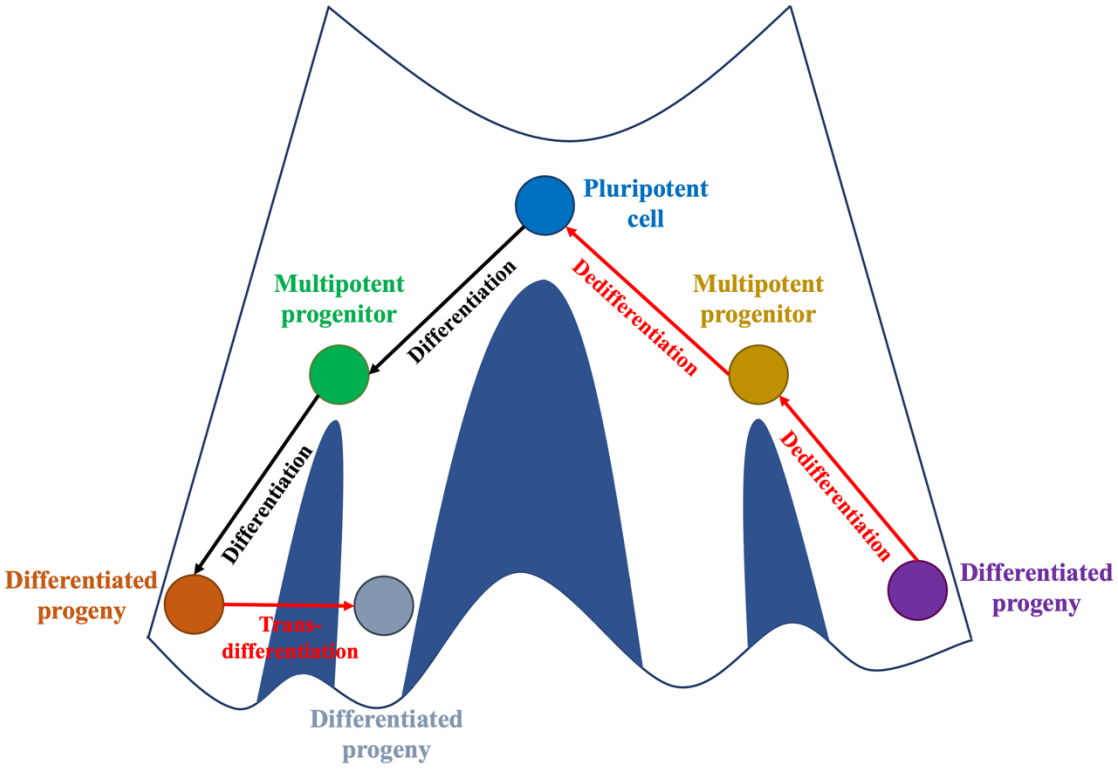


Figure 1.1 A schematic representation of the Waddington's epigenetic landscape concept
The epigenetic landscape idea is visualised by a ball (representing a pluripotent cell) rolling down a slope and encountering multiple branches down which it can choose to travel (the lineage specification choices represented by the different ball colours). In this model, the cell cannot spontaneously move back up the slope (dedifferentiate) or move across to another path (transdifferentiate), as shown by the red arrows. This is due to certain epigenetic barriers being put in place in its DNA at every differentiation stage.

Together, these factors can alter the chromatin between the permissive (euchromatin) and non-permissive (heterochromatin) states, thus affecting the ability of transcription factors to bind to target DNA sequences and hence regulating gene expression patters (reviewed by Klemm et al., 2019).

The work described in this thesis focuses specifically on the cytosine methylation and hydroxymethylation as well as the family of proteins responsible for generation of the latter chemical group. The role of these epigenetic DNA modifications is studied in the processes of cancer metastasis and cellular differentiation.

1.2 The power of the methyl group - cytosine methylation in physiology and disease

1.2.1 The discovery and mechanisms of DNA methylation

The methylation of a cytosine base at carbon position 5, also termed the "fifth DNA base", is the most well studied and understood epigenetic mark (Stricker and Götz, 2018). Cytosine methylation in mammals was discovered only a few years after DNA was described as the source of cells' genetic information (Avery et al., 1944; Hotchkiss, 1948). Rollin Hotchkiss utilised paper chromatography to separate the DNA bases and observed a base migrating at rate slightly higher than that of cytosine, and which had a considerably lower abundance than cytosine. He termed this modification "epicytosine" and even suggested that it might be 5-methylcytosine based on the previous description of this base in *Tubercle bacillus* (Johnson and Coghill, 1925; Hotchkiss, 1948). Three decades after this discovery, 5mC was proposed to play a role in regulating gene expression, although the exact mechanism was still unclear (Holliday and Pugh, 1975). In the 1980s, it was already known that methylation patterns are both heritable and tissue-specific (Razin and Riggs, 1980). The 5mC mark was also observed to occur in the context of CpGs, which denote a cytosine followed immediately by a guanine, and CpG islands (CGIs) which is name given to a number of CpGs clustered in close proximity within the DNA sequence (Bird, 1980; Bird et al., 1985). The CpGs are now known to be between 70 to 80% methylated, although methylation can also exist in the non-CpG context (Jang et al., 2017).

The enzymes responsible for methylating cytosines were discovered when the mouse DNA methyltransferase 1 (DNMT1) protein was described and sequenced by Timothy Bestor and colleagues (Bestor and Ingram, 1983; Bestor et al., 1988). DNMT1 was found by the authors to have a preference for hemimethylated DNA, which refers to

only one of the DNA strands being methylated. It is now known that this is because DNMT1 is predominantly a maintenance methyltransferase responsible for copying the methylation pattern of the template strand onto the newly synthesised DNA strand following the replication process (Edwards et al., 2017). DNMT1 was also found to be a self-inhibitory enzyme that requires interaction with the Ubiquitin-like with PHD and ring finger domains 1 (UHRF1) protein for DNA binding and subsequent cytosine methylation (Liu et al., 2013; Li et al., 2018). Subsequent to that, two *de novo* methyltransferases DNMT3A and DNMT3B were discovered by Okano and colleagues (Okano et al., 1998a; Okano et al., 1999). The two enzymes were shown by the authors to have no preference for hemimethylated DNA and to be essential for mammalian embryonic development. Additionally, DNMT2 and DNMT3L have also been identified. DNMT2 was shown to be non-essential in embryonic stem cells for neither *de novo* nor maintenance methylation (Okano et al., 1998b). However, the DNMT2 protein is the most conserved member of the DNMT family in eukaryotes and more recently it has been identified as transfer RNA methyltransferase (Goll et al., 2006; Ashapkin et al., 2016). DNMT3L was identified later than the other two members of the DNMT3 family (Aapola et al., 2000), but did not display any *de novo* methyltransferase activity (Hata et al., 2002). It was instead found to regulate *de novo* methylation by interacting with DNMT3A and DNMT3B and enhancing their catalytic activity, especially during the embryogenesis process (Chédin et al., 2002; Suetake et al., 2004; Gowher et al., 2005; Van Emburgh and Robertson, 2011). The DNMT family establishes the methylation mark on the cytosine bases via a reaction involving a compound called S-adenosyl-L-methionine, also known as SAM, which provides the methyl group that is attached to the carbon position 5 of the unmodified cytosine by the DNMT enzyme with S-adenosyl-L-homocysteine produced as a by-product (Zhang and Zheng, 2016).

1.2.2 The link between inactivation of gene expression and cytosine methylation

The repressive effects of cytosine methylation on gene expression were first proposed in the 1980s (Razin and Riggs, 1980). The authors observed that “undermethylation” was correlated with activation of gene expression in a number of studies. Since then, 5mC was proposed to silence gene transcription by either physically preventing transcription

factors from interacting with DNA or by binding of repressor proteins to the methylated DNA sequences (Tate and Bird, 1993). More recent observations indicate that the direct mechanism of physical occlusion of transcription factor binding is likely to play a role in transcriptional repression caused by 5mC, but it is not the only mechanism responsible for it (Choy et al., 2010; Medvedeva et al., 2014). Methyl binding proteins, such as methyl-CpG binding protein 1 (MeCP1) (Meehan et al., 1989) and MeCP2 (Lewis et al., 1992) have been shown to interact specifically with methylated cytosines and act as mediators of transcriptional silencing (Cross et al., 1997). An example of the mechanism of such gene expression repression involves a complex formed of MeCP2, SIN3A and histone deacetylase proteins: HDAC1 and HDAC2 (Nan et al., 1998). The deacetylation of histones utilises the nuceosomal assembly to subsequently change chromatin confirmation to a repressive state (Jones et al., 1998). More recently, it has also been proposed that the presence of 5mC may create new binding sites for transcription factors, which are not recognised by these proteins when DNA is unmethylated (reviewed by Zhu et al., 2016). The examples of these include KAP1 and ZFP57, which have specific recognition sequences that include methylated cytosines (Quenneville et al., 2011; Liu et al., 2012). Additionally, despite the consensus that 5mC is predominantly associated with repression of transcription, it has been shown that gene body methylation is associated with transcriptional activation for a number of genes (Ball et al., 2009). Thus, the effects of DNA methylation are context dependent.

1.2.3 The fifth DNA base during mammalian development

The 5-methylcytosine base has been found to play critical roles at numerous stages during mammalian development due to its role in transcriptional regulation. The critical importance of DNA methylation in the developmental processes is highlighted by embryonic lethality of *Dnmt1* and *Dnmt3b* knockout mice and the growth retardation and neonatal lethality of the *Dnmt3a* knockout mice (Li et al., 1992; Okano et al., 1999). In general, mammalian development is characterised by two waves of DNA demethylation and subsequent re-methylation. In humans, following fertilisation both parental genomes are demethylated until the blastocyst stage, with the paternal genome showing a more rapid rate of demethylation than the maternal one (Guo et al., 2014). The majority of this genome-wide 5mC loss occurs between the fertilisation and the two

cell stage, which is contrast to the mouse embryonic development where most of demethylation occurs at the one cell stage (Guo et al., 2014; Okamoto et al., 2016). There are genomic regions, however, that resist this first wave of demethylation. Most notably, these include imprinting control regions (ICRs) responsible for regulating the expression of imprinted genes (Quenneville et al., 2011), although there are differentially methylated regions (DMRs) other than ICRs that resist demethylation until the blastocyst stage (Proudhon et al., 2012). Imprinted genes are expressed from either a paternal or maternal allele while being silenced by DNA and histone methylation on corresponding allele inherited from the other parent (reviewed by Reik, 2007). The silencing is maintained through this first wave of global demethylation via recruitment of the KAP1 (KRAB-associated protein 1, also known as TRIM28) by Krüppel-associated box (KRAB) domain containing zinc-finger proteins (ZFPs) (Quenneville et al., 2011; Messerschmidt et al., 2012). KAP1 in turn acts as a scaffold for a number of proteins including HP1 and histone methyltransferases (HMTs) and HDACs to establish heterochromatin confirmation around the region that protects it from the loss of 5mC mark (Schultz et al., 2002; Alexander et al., 2015). Lastly, KAP1 and KRAB-ZFPs are also thought to play a role in maintenance of cytosine methylation at transposable elements (Rowe et al., 2010; Coluccio et al., 2018; Stoll et al., 2019). These genomic regions contain the majority of 5mC persisting up to the blastocyst stage and in particular include the evolutionarily younger LINE-1 and SVA elements in human embryos (Guo et al., 2014) while in mice the IAP endogenous retrovirus sequences are particularly resistant to methylation loss (Rowe et al., 2010; Smith et al., 2012).

The second wave of demethylation during mammalian development occurs during the formation of primordial germ cells (PGCs). These are precursors to the sperm and the eggs and hence need to acquire the sex-specific imprints that will be passed on the next generation following fertilisation. This is achieved by removing methylation from the ICRs that were protected during the first demethylation wave (Hargan-Calvopina et al., 2016). The mouse PGCs are first specified at embryonic day 6.5 (ED6.25) from epiblast cells, detected at approximately ED7.25 and migrate to the genital ridges by ED10.5 where the female germ cell enter meiosis at ED13.5 while the male germ cells experience mitotic arrest (Lawson et al., 1999; Saitou and Yamaji, 2012). In humans this period is considerably longer and lasts between approximately week 3 and week 10

post fertilisation, although this has not been extensively studied due to ethical limitations (Leitch, et al., 2013). More recent evidence, however, suggest that human PGC specification occurs earlier than previously thought, at around day 12 post fertilisation (Chen et al., 2019).

The demethylation of mammalian PGCs is thought to occur in two waves (Seisenberger et al., 2012; Kobayashi et al., 2013). The first wave of DNA demethylation occurs from approximately ED8 and involves passive loss of the 5mC mark through downregulation of *de novo* DNMT3A and DNMT3B methyltransferases as well as UHRF1 (Kagiwada et al., 2013). This results in cytosine methylation not being maintained and deposited in the genome through successive rounds of PGC replication. The second DNA demethylation wave begins around ED9.5 and involves active DNA demethylation by the ten-eleven translocation enzymes (TETs) resulting in the lowest levels of 5mC in PGCs of both sexes at ED13.5 (Seisenberger et al., 2012; Dawlaty et al., 2013; Hackett, et al., 2013). For a comprehensive description of the role of TETs in reducing the global methylation levels in PGCs, see section 1.4.1. The two gametes are subsequently remethylated, which establishes the correct, sex-specific genomic imprints. In sperm this occurs after ED13.5 while oocytes only complete the process after birth (reviewed by Kota and Feil, 2010; Hanna and Kelsey, 2014).

1.2.4 The role of cytosine methylation in cellular differentiation - embryonic stem cells and the small intestine

The process of cellular differentiation involves the progressive establishment of cellular identity from the pluripotent stem cell state. The unique characteristics of a terminally differentiated cell are maintained through epigenetic mechanisms, including cytosine methylation (reviewed by Khavari et al., 2010). The mouse embryonic stem cells (mESC) show low levels of CpG methylation (around 30%) compared to the somatic tissues (approximately 80%) thus enabling the plasticity of ESCs to differentiate into a chosen lineage upon an appropriate stimulus while retaining methylation at selected regions such as retrotransposons and imprinted genes (Hon et al., 2013; Lee et al., 2014). This is thought to be established by reduced levels of DNMT3A, DNMT3B and DNMT3L as well as the activity of TET1 and TET2 (see section 1.4.2) (Ficz et al., 2013; Hackett, et al., 2013). When ESCs begin to differentiate, the opposite pattern is

observed with a drop in TET1 and TET2 levels coupled with an increased activity of the DNMT enzymes (Oda et al., 2013). This establishes methylation marks in the pluripotency-associated genes, such as *Nanog* or *Lefty1*, to silence them as the cells exit the pluripotent state (Farthing et al., 2008). The global gain in DNA methylation observed during differentiation is also involved in silencing the genes non-specific to the chosen lineage (Calvanese et al., 2012); however there are also genes which are specific to the differentiated progeny that loose methylation in order to be expressed (Liu et al., 2017; Cypris et al., 2019). This is achieved in part through the oxidising activity of the TET enzymes (see section 1.4.4).

The mouse small intestine represents an excellent model for studying adult stem cell differentiation due to well-defined cellular subpopulations and rapid turnover of cells (Forn et al., 2015; Jorgensen and Ro, 2019). The structure of the small intestine has been extensively studied and a number of cell types were identified (Figure 1.2). The intestinal stem cells (ISCs) residing in the crypt are defined by their *Lgr5* expression, self-renewal capabilities and the ability to differentiate into all the cell types making up the mouse small intestine (Barker et al., 2007). Within the intestinal crypt, the cells are numbered based on their position relative to the bottom of the crypt. The cells at position "+4" are thought to be a "reserve" stem cell, although this hypothesis and the true identity of these cells are still controversial (Barker et al., 2012).

The other cell type residing in the intestinal crypt is the Paneth cell which provides the niche for the ISCs through factors and signals required for their maintenance (Meran et al., 2017). The ICSs give rise to the transit amplifying (TA) progenitors, which move through the TA zone to the villus to become one of the fully differentiated cell types residing there. These include the enterocytes, which represent the majority of the cells present in the intestinal villi, tuft cells, goblet cells and enteroendocrine cells with arterioles and venules responsible for blood supply and lacteals involved in fat absorption (Meran et al., 2017). In mice, the process of differentiation and migration of cells from the crypt niche to the tip of the villus is typically completed within 3-5 days thus providing a rapid turnover of cells that is advantageous for studying the differentiation process (Barker et al., 2007).

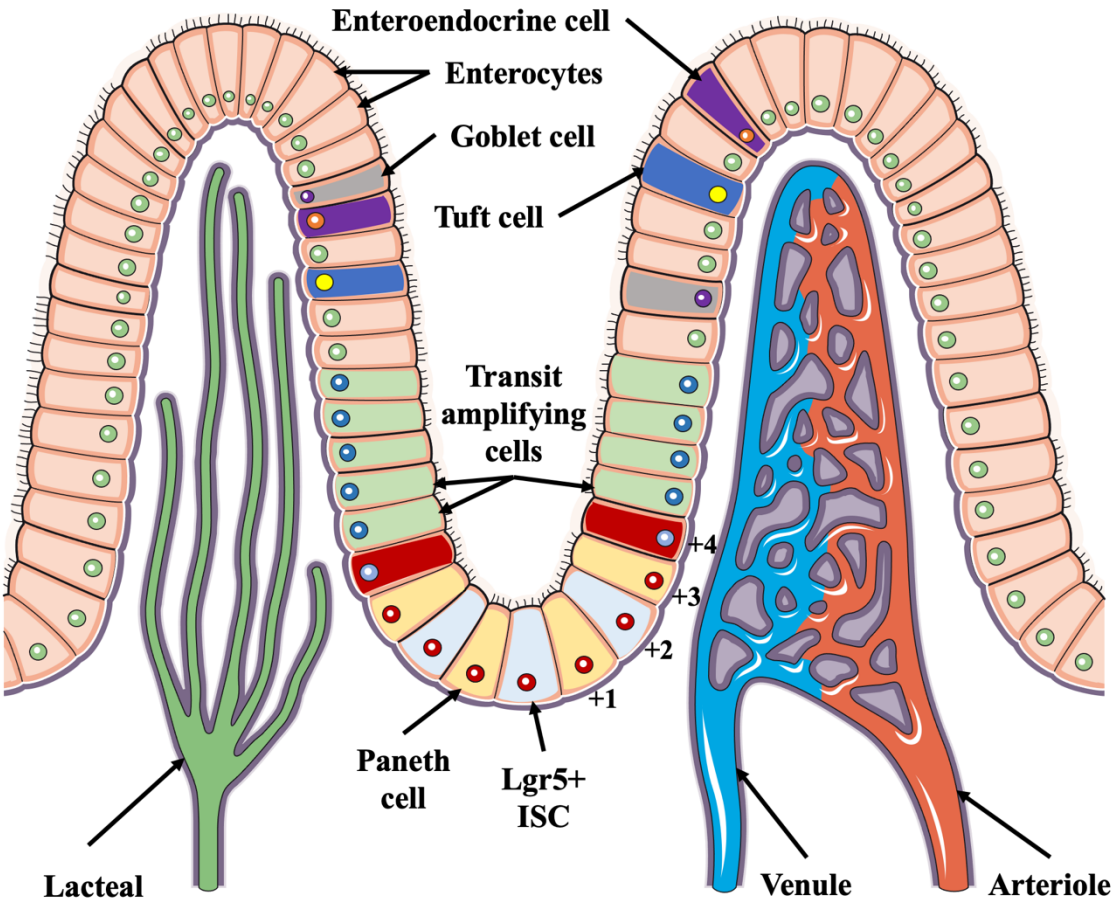


Figure 1.2 The schematic structure of the epithelium lining the small intestine
The schematic representation of the epithelium forming the crypt and the villi in the small intestine. The numbers represent the positions of the cells relative to the base of the intestinal crypt. Lgr5+ ISC = Lgr5-positive intestinal stem cell.

The crypts are characterised by more rapid proliferation rate that the villi, as measured by the Ki67 marker expression while the global 5mC levels remain largely unchanged along the crypt-villus axis (Kaaij et al., 2013; Uribe-Lewis et al., 2020). This is in contrast to the global 5hmC levels, which are higher in the differentiated villus cells than in the crypt stem cell niche (see section 1.4.2) (Kim et al., 2016; Uribe-Lewis et al., 2020). The analysis of the correlation between local DNA methylation changes by Sheaffer and colleagues revealed that 5mC plays an important role in regulation of the differentiation process (Sheaffer et al., 2014). Through combination of whole genome methylation and transcript analysis, the authors showed that gain in 5mC levels during differentiation is associated with silencing the genes required for stem cell function, such as *Lgr5* and *Olfm4*, while enterocyte-specific genes such as *AP* or *Lct* showed reduced methylation. The vast majority of these changes were found to occur in the

introns of these genes. Additionally, numerous DMRs between the stem and differentiated cells are located in enhancers, whose demethylation contributes to activation of gene expression by facilitation of transcription factors binding (Sheaffer et al., 2014).

There is some controversy regarding the extent of methylation changes occurring during the intestinal differentiation with Kaaij and colleagues reporting a modest 50 DMRs across the whole genome (Kaaij et al., 2013). However, this discrepancy is likely to have resulted from the strict DMR definition adopted by the authors, with DMRs specified by at least 40% difference in 5mC levels across at least 10 CpGs. Lastly, the importance of DNMT1 in the differentiation process was highlighted by Sheaffer and colleagues who used a conditional mouse *Dnmt1* knockout to demonstrate an expansion of the intestinal crypt and reduction of differentiation markers in the villus epithelium (Sheaffer et al., 2014). Taken together, these results point towards a critical role of 5mC in both embryonic and adult stem cell differentiation.

1.2.5 The disease states associated with 5-methylcytosine

The evidence of critical role of the 5mC mark in numerous physiological processes described above makes it perhaps unsurprising that its deregulation is implicated in a number of human diseases. While the detailed description of all these conditions is beyond the scope of thesis, the examples discussed below give insight into the variety of diseases associated with perturbations in 5mC deposition and maintenance. For a detailed description of DNA methylation's role in cancer, see section 1.2.6.

A number of rare human diseases arise due to dysregulation of the imprinting process. Beckwith–Wiedemann syndrome occurs in approximately 1 in 13,000 births and is caused by pathological alterations to maternal chromosome region 11p15.5 which contains *CDKN1C* and *IGF2* imprinted genes that play an important role in growth regulation of the foetus (Brioude et al., 2018). This disease is mainly diagnosed in children and is characterised by tongue and organ overgrowth and predisposition to embryonic tumours, most notably Wilms' kidney tumour (Weksberg et al., 2001). Prader-Willi and Angelman syndromes are diagnosed in 1 in 20,000 and 1 in 15,000 births, respectively, and both arise from imprinting defects in the 15q11-q13

chromosomal region (Nicholls and Knepper, 2001; Goldstone, 2004). Prader-Willi syndrome results from a deletion of the abovementioned chromosomal region on the paternal allele while the same deletion, but on the maternal allele, causes the Angelman syndrome (Kim et al., 2019). Both syndromes present with mental retardation and abnormal behavioural patterns, with the Prader-Willi sufferers also developing hyperphagia and obesity and Angelman syndrome patients displaying speech impairment and ataxia (Goldstone, 2004; Buiting et al., 2016).

Beyond dysregulation of genomic imprinting, DNA methylation has also been implicated in autoimmune disorders, although the evidence is not as strong as for the imprinting disorders. These include rheumatoid arthritis (Liu et al., 2013; Ciechomska et al., 2019), systemic lupus erythematosus (Wu et al., 2016; Hedrich et al., 2017) and multiple sclerosis (Chomyk et al., 2017; Li et al., 2017; Maltby et al., 2018; Souren et al., 2019). Multiple X chromosome-linked disorders presenting with mental retardation were also shown to be associated with aberrant DNA methylation. These include Fragile X syndrome (Oberlé et al., 1991; Schenkel et al., 2016) and alpha-thalassemia mental retardation syndrome (Gibbons et al., 2000; Schenkel et al., 2017). Lastly, DNA methylation is known to be closely linked with the process of aging, which was recently proposed to be classified as a disease (Bulterijs et al., 2015; Zhavoronkov and Bhullar, 2015; Lustgarten, 2016; Zhavoronkov and Moskalev, 2016; Khaltourina et al., 2020). This link was famously established by Steve Horvath's research group, who proposed an epigenetic clock estimating the biological age of cells, tissues and individuals (Horvath, 2013; Horvath, 2015; Levine, et al., 2015; Marioni et al., 2015; Christiansen et al., 2016). This aging clock is based on the methylation status of 353 CpG sites across the genome (Horvath, 2013), and higher epigenetic age measured by Horvath's method was found in, amongst others, Alzheimer's and Parkinson's disease sufferers (Horvath and Ritz, 2015; Levine et al., 2015).

1.2.6 The dark side of the fifth DNA base - how cytosine methylation contributes to tumorigenesis

The first indication of the role of DNA methylation in cancer pathophysiology came from Andrew Feinberg and Bert Vogelstein who demonstrated a reduction in methylation levels of a substantial number of CpGs in cancer cells (Feinberg and

Vogelstein, 1983a). The lower global 5mC levels in cancer cells compared to normal tissue were subsequently confirmed by Gama-Sosa and colleagues, with metastatic tumours showing a reduced DNA methylation content than benign ones (Gama-Sosa et al., 1983). This global DNA hypomethylation results in activation in numerous oncogenes, including *RAS* genes (Feinberg and Vogelstein, 1983b), metastasis-associated *S100A4* gene (Nakamura and Takenaga, 1998), *HPV16* genome (Badal et al., 2003) and *CCND2* gene (Oshimo et al., 2003). Additionally, global loss of 5mC leads to increased genomic instability and genomic rearrangements which contribute to the tumorigenesis process (Qu et al., 1999; Sheaffer et al., 2016). Lastly, it was found that loss of genomic imprinting contributes to the tumorigenesis process, most notably in the embryonic Wilms' tumours (Steenman et al., 1994; Mummert et al., 2005) and chronic myeloid leukaemia (Randhawa et al., 1998; Benetatos and Vartholomatos, 2015). A well-studied example of this is the first discovered imprinted genes, the *IGF2* and *H19* locus, whose dysregulation of imprinting has been implicated in a variety of cancers (Hibi et al., 1996; Cui et al., 2002; Byun et al., 2007; Zhao et al., 2016; Gailhouste et al., 2019).

The reduced global levels of 5mC in cancer cells do not correspond to the changes in DNA methylation content in numerous genomic loci, most notably within tumour suppressor genes. This was initially demonstrated through hypermethylation of the *RB* gene, the first tumour suppressor gene to be discovered, leading to the loss of its expression (Greger et al., 1989; Sakai et al., 1991; Ohtani-Fujita et al., 1993; Greger et al., 1994). Since then, hypermethylation of tumour suppressor genes was demonstrated in a plethora of cancer types, notably in the *Tp53* tumour suppressor gene described as the "guardian of the genome" whose importance in prevention of tumorigenesis is highlighted by the fact that it is mutated in approximately 50% of human cancers (Aubrey et al., 2018). The *Tp53* promoter was shown to be methylated in numerous cancers, including chronic lymphocytic leukaemia (Saeed et al., 2019), ovarian cancer (Chmelarova et al., 2013) and breast cancer (Kang et al., 2001). Novel examples of tumour suppressor gene hypermethylation continue to be identified, such as recent descriptions of promoter methylation of: *TET1* in numerous cancer types (Li et al., 2016), *ECRG4* in breast cancer (Tang et al., 2019), *CDH11* in colorectal cancer (Yuan et al., 2019) and *CDO1* in lung cancer (Yin et al., 2020). In fact, a distinct term was proposed by Toyota and co-authors in 1999 for cancers characterised by promoter

hypermethylation of a number of tumour suppressor genes simultaneously: CpG island methylator phenotype (CIMP) (Toyota et al., 1999). In particular, colorectal cancer is characterised by high incidence of this phenotype, with approximately 20% of colorectal tumours assigned CIMP, which was shown to be associated with microsatellite instability, older age and poor tumour differentiation (Jia et al., 2016; Advani et al., 2018; Advani et al., 2019).

1.3 Passive and active DNA demethylation - the discovery of the TET enzymes and 5hmC

1.3.1 Another brick in the epigenetic wall - the brief history of cytosine hydroxymethylation

The existence of the 5hmC base in the DNA has been known for close to seventy years, when its presence was confirmed in bacteriophage DNA (Wyatt and Cohen, 1952; Hershey et al., 1953). Approximately twenty years later, it was also discovered in mammalian DNA, although the 15% of cytosine levels reported by Penn and colleagues in rat brain and liver are now known to be in fact fifteen-fold lower (Penn et al., 1972; Globisch et al., 2010). In 2009, two groups independently reported the presence of 5hmC in murine Purkinje neurons (Kriaucionis and Heintz, 2009) and embryonic stem cells (Tahiliani et al., 2009) by thin layer chromatography and mass spectrometry. These two publications have subsequently created a new field within epigenetics with 5hmC now known to play critical roles during mammalian development, cellular differentiation and tumorigenesis.

1.3.2 TET1 and the discovery of the TET enzymes

The laboratory group of Anjana Rao, responsible for the initial description of 5hmC in mouse ESCs, was the first one to identify the protein responsible for its deposition, Ten-eleven translocation 1 (TET1), as a fusion partner of the *MLL* gene in acute myeloid leukaemia harbouring the t(10;11)(q22;q23) chromosomal translocation (Tahiliani et al., 2009). The authors discovered this enzyme, along with TET2 and TET3, through a

homology search using the trypanosome J binding protein 1 and 2 as a reference. Through TET1 overexpression and the use of mass spectrometry, they confirmed that TET1 was indeed responsible for conversion of 5mC to 5hmC in an oxygen, iron (II) and α-ketoglutarate dependent manner (Tahiliani et al., 2009). Subsequently, the other two TET proteins, TET2 and TET3, were shown to be capable of generating 5hmC in the same way (Ito et al., 2010). The TET enzymes were demonstrated to all contain a conserved C-terminal region comprised of a cysteine-rich region and a double stranded beta-helix domain, where the iron (II) and α-ketoglutarate binding regions are located (Ito et al., 2010; Yin and Xu, 2016). While TET1 and TET3 contain a CXXC domain responsible for binding the DNA in both unmethylated and methylated CpG context, but the *TET2* gene does not encode such a domain (reviewed by Yin and Xu, 2016). The CXXC domain of TET2 is in fact encoded by a separate gene called *IDAX*, which also regulates *TET2* expression (Ko et al., 2013). Isoforms of TET1 and TET3 which do not contain the CXXC domain but are still recruited to DNA, albeit in TET1's case to a reduced degree, by other proteins have also been described (Liu et al., 2013; Perera et al., 2015; Zhang et al., 2016; Good et al., 2017). In particular, the truncated TET1 isoform was proposed to be overexpressed in glioblastoma, breast and uterine cancers with the latter two characterised by an association between high TET1 levels and lower overall survival (Good et al., 2017).

1.3.3 Passive and active DNA demethylation

The process of removal of methyl groups from cytosines can occur either passively or actively. The passive DNA demethylation occurs through a downregulation in the levels or activity of DNMT1 and its cofactor UHRF1, leading to a reduction in the maintenance of methyl marks during successive rounds of cell proliferation (Wang et al., 2009; He et al., 2017; Seo et al., 2017). The evidence from work done in ESCs also supports the role of reduced levels of *de novo* methyltransferases DNMT3A and DNMT3B, as well as DNMT3L, in the passive demethylation process (Ficz et al., 2013; Leitch et al., 2013; Yamaji et al., 2013).

Active DNA demethylation involves a number of enzymatic reactions, the most prominent and widely accepted of which is the iterative oxidation of the 5mC base by the TET proteins followed by base excision repair (BER) through the thymine DNA

glycosylase (TDG) enzyme (Figure 1.3) (reviewed by Wu and Zhang, 2017). Both 5fC and 5caC, but not 5hmC, were shown to be recognised and excised by TDG (He et al., 2011; Maiti and Drohat, 2011). The resulting abasic site is then incised by the APE1, NEIL1 or NEIL2 endonucleases removing the baseless sugar moiety, followed by the insertion of an unmodified cytosine by DNA polymerase β and DNA ligation trough the combined action of XRCC1 and LIG3 proteins thus completing the demethylation process (Schomacher et al., 2016; Weber et al., 2016).

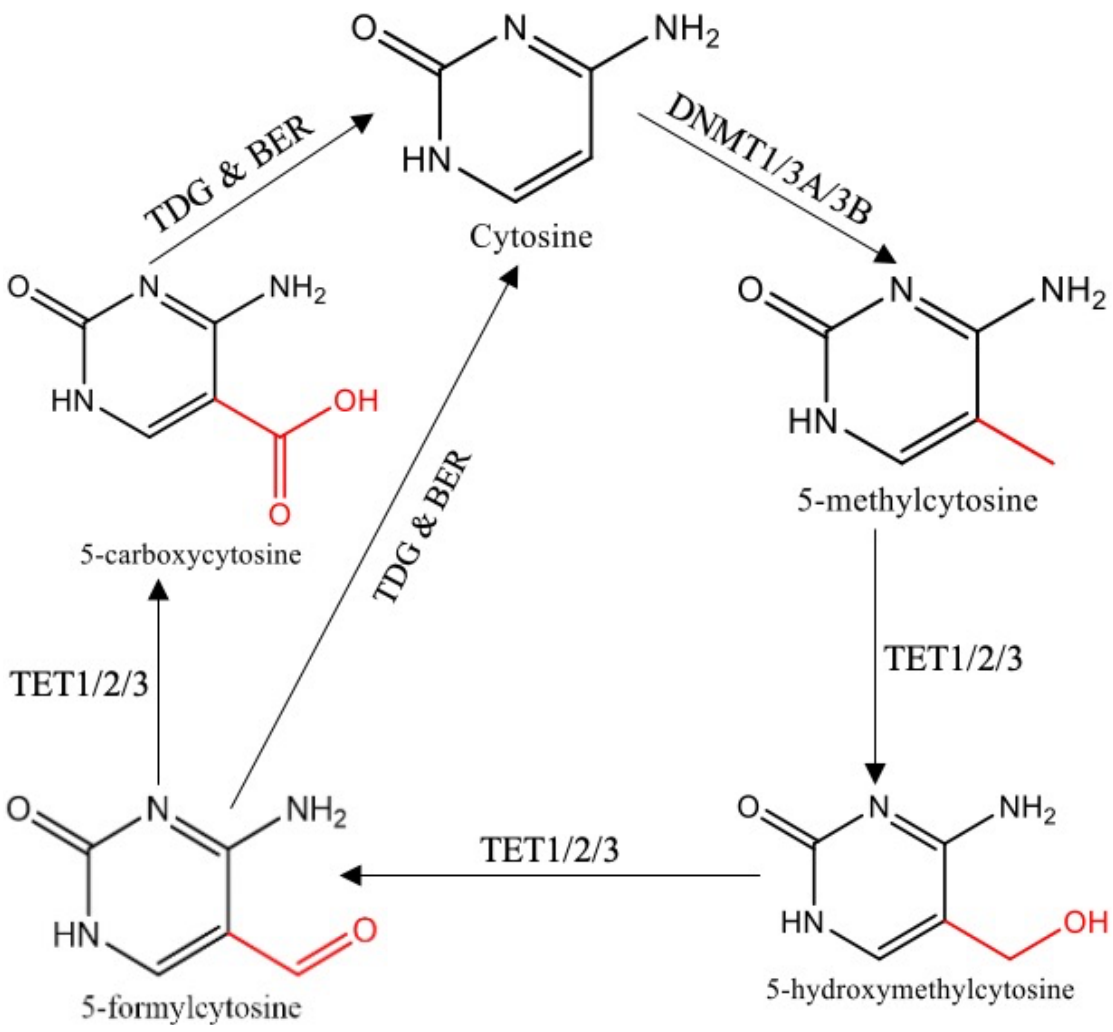


Figure 1.3 The active demethylation of cytosine mediated by the TET and TDG enzymes

The schematic structural representation of the demethylation cycle of the cytosine base. The chemical group modified by the enzymatic reactions is highlighted in red. DNMT = DNA methyltransferase; TET = ten-eleven translocation; TDG = thymine DNA glycosylase; BER = base excision repair.

The removal of the methyl group from cytosine bases can also occur in a semi-passive manner. This process involves the oxidation of 5mC by the TET enzymes followed by passive removal of the resulting 5hmC, 5fC or 5caC bases through successive rounds of DNA replication. This is possible due to reduced preference of the DNMT1 enzyme for the oxidised forms of 5mC, which in turn can result in dilution of these bases from the DNA during cell proliferation (Hashimoto et al., 2012; Ji et al., 2014). Additionally, UHRF1 was suggested to have a reduced affinity towards 5hmC as compared to 5mC, which may therefore result in less efficient recruitment of DNMT1 to 5hmC-containing CpGs compared to 5mC-containing ones (Hashimoto et al., 2012; Otani et al., 2013).

The activation-induced deaminase and apolipoprotein B mRNA editing enzyme, catalytic polypeptide (AID/APOBEC) family of enzymes was also implicated in the active DNA demethylation process. This was based on evidence from *in vitro* and E.coli experiments (Morgan et al., 2004) and *in vivo* zebrafish work (Rai et al., 2008), with the zebrafish AID enzyme displaying a unique affinity amongst its mammalian orthologues towards deaminating 5mC to thymine (Abdouni et al., 2013). In the latter scenario, AID was proposed to cooperate with the MBD4 glycosylase and the GADD45A DNA repair protein to remove the thymine-guanine mismatch and complete the demethylation cycle (Rai et al., 2008). However, a contradictory report failed to observe a deamination activity of AID on 5mC in a double-stranded DNA context (Bransteitter et al., 2003). Additionally, the AID/APOBEC proteins were shown to have little *in vitro* activity on 5mC, and no activity on 5hmC, as compared to an unmodified cytosine (Nabel et al., 2012). Lastly, AID-knockout mice do not exhibit severe developmental defects aside from immune system issues and lymphoma formation (Muramatsu et al., 2000; Kotani et al., 2007) and APOBEC1-knockout mice appear healthy and fertile (Hirano et al., 1996; Morrison et al., 1996), thus casting further doubt on the importance of these proteins in the DNA demethylation process.

1.3.4 Are 5-hydroxymethylcytosine and 5-formylcytosine independent epigenetic marks?

The initial description of the TET enzymes' ability to convert 5mC to 5hmC (Ito et al., 2010) as well as 5fC and 5caC (Ito et al., 2011) identified these bases as intermediates in the active DNA demethylation pathway. However, studies from the Murrell and

Balasubramanian groups provided evidence that 5hmC is a mostly stable DNA modification and 5fC can be stable, especially in the brain (Bachman et al., 2014; Bachman et al., 2015). A number of 5hmC and 5fC-binding proteins were identified including MeCP2 (Mellén et al., 2012), MBD3 (Yildirim et al., 2011) and UHRF1/2 (Spruijt et al., 2013; Chen et al., 2017). Additionally, a screen of binding proteins of these two cytosine modifications found more proteins with specificity towards 5fC than 5hmC, including TDG and a number of the FOX transcription factors (Iurlaro et al., 2013). The 5fC base was found to alter the structure of the DNA double helix by causing its underwinding (Raiber et al., 2015). Taken together, this evidence suggests that both 5hmC and 5fC are epigenetic modifications in their own right.

1.3.5 The wind of change - the emerging roles of cytosine hydroxymethylation and 5-formylcytosine beyond active demethylation

The distribution of the 5hmC base in the genome strongly suggests an association with active gene transcription (reviewed by Lio and Rao, 2019). Firstly, the presence of 5hmC within gene bodies is positively correlated with gene expression with 5hmC found to be present in the gene bodies of highly transcribed genes in multiple tissues and cell types, including the brain (Wang et al., 2012; Lister et al., 2013), T-cells (Tsagaratou et al., 2014), ESCs (Pastor et al., 2011), liver (Ivanov et al., 2013) and small intestine (Uribe-Lewis et al., 2020). The ESC bivalent promoters, characterised by the presence of both the repressive H3K27me3 and activating H3K4me3 histone tail marks, mark genes whose transcription is only initiated during cellular differentiation (Blanco et al., 2020). These promoters are known to harbour 5hmC to protect them from hypermethylation which would otherwise stop their expression during differentiation (Pastor et al., 2011; Verma et al., 2018).

The 5hmC base is also well documented to accumulate at the distal regulatory enhancer regions responsible for regulation of gene transcription levels (Mahé et al., 2017; Shi et al., 2017). The deposition 5hmC mark is associated with enhancer activation and this process was demonstrated to be mediated by the TET2 enzyme (Sérandour et al., 2012; Hon et al., 2014; Yang et al., 2016; Wang et al., 2018). The presence of 5hmC within exons and exon-intron boundaries was suggested to play a role in alternative splicing (Shi et al., 2017). It was found to promote the inclusion of the exon in the transcript in a

CTCF-mediated mechanisms in T-cells and MeCP2-dependent way in the brain, with no such function observed in the liver indicating the tissue specificity of this mechanism (Khare et al., 2012; Marina et al., 2016; Cheng et al., 2017). Lastly, 5hmC was shown to colocalize with DNA repair proteins to DNA damage loci in HeLa cells in a TET2-dependent manner (Kafer et al., 2016). The authors also described evidence for the role of TET enzymes in genome integrity maintenance. A separate report confirmed the 5hmC accumulation at DNA damage foci in numerous cell types and demonstrated the contribution of the TET3 enzyme to ATR-mediated DNA repair (Jiang et al., 2017). While the role of the TET enzymes and 5hmC in DNA repair is far from established and further work is required to elucidate it, it is tempting to speculate that the reduced levels of both in cancer cells could be contributing to the tumours' genomic instability.

The 5fC nucleotide is present in the genome at 0.00002 to 0.0011% of cytosine content with the base found to be enriched at active enhancers in a tissue-specific manner in the mouse embryo (Iurlaro et al., 2016). It was shown to induce a structural change in the DNA resulting in helical underwinding (Raiber et al., 2015), although conflicting results published by the Brown group mean that this effect is still debatable (Hardwick et al., 2017). More recent report suggested that 5fC creates a localised change to the DNA structure which is recognised by the TDG enzyme prior to the base excision (Fu et al., 2019). The presence of 5fC in the promoter region was demonstrated to repress gene expression (Kitsera et al., 2017) and it is also known to reduce the rate of transcription by the RNA polymerase II (Kellinger et al., 2012). Lastly, 5fC is known to induce DNA-protein crosslinks with the histone proteins (Li et al., 2017; Ji et al., 2019). Together, these findings suggest that the 5fC base may play a role in transcriptional control and nucleosomal regulation.

1.3.6 The mysterious 5-carboxycytosine

The 5caC base still remains poorly understood with its prevalence in mESCs recorded at 0.00005% of cytosine levels (Hofer et al., 2019). In the same fashion as 5fC, it was demonstrated to both stall the RNA polymerase II during transcriptional elongation (Kellinger et al., 2012; Wang et al., 2015) and repress transcription when present within the promoter region (Kitsera et al., 2017). On the other hand, the evidence from mESCs studies suggests that 5caC marks active promoters and enhancer regions (Lu et al.,

2015; Neri et al., 2015). The 5caC base was also found to be enriched in certain brain cancer cell lines (Ramsawhook et al., 2017) as well as breast cancer (Guo et al., 2017) and glioma samples (Eleftheriou et al., 2015), although the possible role of 5caC in tumorigenesis is yet to be elucidated. Lastly, a number of 5caC reader proteins were identified, including TDG (Zhang et al., 2012), the TCF4 transcription factor (Yang et al., 2019) and a specific domain within the TET3 enzyme (Jin et al., 2016). Overall, further work is required to confirm the potential functions of the 5caC base.

1.3.7 The activity of TET enzymes is post-translationally regulated

The TET proteins are known to have multiple sites and regions that can be post-translationally modified (Liu et al., 2019). All three TET enzymes can be modified with O-linked N-Acetylglucosamine (O-GlcNAc) residues by the O-GlcNAc transferase (OGT) enzyme (Bauer et al., 2015). The TET1 enzyme was demonstrated to both interact with OGT and be O-GlcNAcylated in mESCs, which resulted in increased repression of the target genes (Shi et al., 2013; Vella et al., 2013). TET1 was also shown to be modified by OGT *in vitro*, which resulted in the increased activity of the former (Hrit et al., 2018). Conversely, two other reports found that OGT associates with both TET2 and TET3, but not TET1 and induces transcriptional activation (Chen et al., 2013; Deplus et al., 2013). TET3 was also independently shown to interact with OGT, be O-GlcNAcylated and enhance OGT recruitment to DNA (Ito et al., 2014). This discrepancy in findings was suggested to stem from differing experimental approaches (Balasubramani and Rao, 2013), but this is unlikely to account for all the differences in results described above. It is therefore clear that further work is required to fully elucidate the nature of the relationship between the TET proteins and OGT as well as its functional importance.

The TET1 enzyme was demonstrated to be polyADP-ribosylated by the polyADP-ribose polymerase (PARP1) enzyme in both covalent and non-covalent manner which results in activation and suppression of TET1 activity, respectively (Ciccarone et al., 2015). PARP1 is also known to promote *TET1* expression through low promoter methylation and high H3K4 histone trimethylation levels (Ciccarone et al., 2014). The TET2 enzyme was shown to be acetylated by the p300 acetyltransferase which

stabilizes it and promotes its activity, and was suggested to protect cells from aberrant methylation due to oxidative stress (Zhang et al., 2017). The phosphorylation of TET2 by the JAK2, AMPK and CDK5 kinases also results in its activation and increased stability (Wu et al., 2018; Jeong et al., 2019; Rao et al., 2020). The site of TET2 phosphorylation by AMPK is bound by the 14-3-3 proteins which protects it from dephosphorylation; a process that was suggested to be deregulated in cancer via TET2 mutations (Chen et al., 2019; Kundu et al., 2020). The DCAF1-containing E3 ubiquitin ligase complex monoubiquitylates the TET proteins promoting their binding to the DNA (Nakagawa et al., 2015). All three TET enzymes are also known to be degraded by the calpain proteases (Wang and Zhang, 2014).

Different miRNAs have been demonstrated to regulate the TET enzymes (Fu et al., 2013; Zhang et al., 2013). TET1 is regulated by miRNAs both during cellular differentiation and in cancer cells. The miRNA-191 was shown to play an oncogenic role in liver and endometrial cancers through downregulation of *TET1* expression (Li et al., 2017; Yang et al., 2020). In hepatocellular carcinoma both miRNA-494 and miRNA-520b suppress tumour progression through reduction of *TET1* level (Chuang et al., 2015; Zhang et al., 2015). Conversely, miRNA-21-mediated inhibition of *TET1* expression in colorectal cancer resulted in a higher cell proliferation rate (Ma et al., 2018). The miRNA-29 family was shown to repress *TET1* expression during differentiation of mESCs to embryonic bodies (Tu et al., 2015; Cui et al., 2016).

TET2 transcript-targeting miRNAs including miRNA-101, miRNA-29b and c, miRNA-7 and miRNA-125b were demonstrated to promote malignant haematopoiesis where TET2 plays a tumour suppressive role (Cheng et al., 2013). A different group reported a similar effect of miRNA-22, including induction of haematological malignancies in mice and defective differentiation of hematopoietic stem cells (Song et al., 2013). The miRNA-767-5p was found to promote thyroid cancer growth and invasion through *TET2* inhibition (Jia et al., 2020). The TET3 mRNA was demonstrated to be regulated by miRNA-29b in the mouse hippocampus (Kremer et al., 2018). Additionally, TET3 is repressed during mouse cortical development and differentiation of neural progenitor cells (Lv et al., 2014). Lastly, all three *TET* transcripts were demonstrated to be downregulated by miRNA-29a in hepatocellular carcinoma cells which contributed to cancer progression (Chen et al., 2017).

In summary, the activity and functions of the TET proteins can be regulated by multiple mechanisms beyond the availability of oxygen, iron (II) and 2-oxoglutarate. However, more research is clearly needed to fully elucidate the functional importance of these mechanisms and resolve the controversies outlined above.

1.4 The sixth DNA base and its roles in development and cellular differentiation

1.4.1 The TET enzymes as novel regulators of the early mammalian development

The TET enzymes play a critical role in the two major demethylation events occurring during mammalian development in the gametes and PGCs (see section 1.2.3 for details). The *in vivo* studies of the effects of *Tet* knockouts provide insight into the importance of these enzymes in the early mammalian development. While *Tet1* and *Tet2* knockout mice were reported to have no clear developmental impairments (Dawlaty et al., 2011; Li et al., 2011; Ito et al., 2019), homozygous knockout of the *Tet3* gene leads to complete neonatal lethality (Gu et al., 2011). Intriguingly, two separate studies reported partial (Kang et al., 2015) and complete (Khoueiry et al., 2017) embryonic lethality of the *Tet1* knockout mice. It is likely that this discrepancy is due to alternative exon targeting strategies and differences between mouse strains used (Khoueiry et al., 2017). The majority of *Tet1* and *Tet2* double knockout mice die perinatally and display developmental defects, but the viable animals are fertile, albeit with reduced female fertility (Dawlaty et al., 2013). This points to the fact that even the double knockout of *Tet1* and *Tet2* is partially conducive with embryonic development. A double *Tet1* and *Tet3* knockout leads to complete embryonic lethality (Kang et al., 2015). The triple *Tet* knockout mice display embryonic lethality with numerous severe defects observed during the gastrulation stage (Dai et al., 2016; Li et al., 2016). These experiments highlight the critical importance of the TET enzymes, and TET3 in particular, in the early mammalian development. The summary of the developmental effects of *Tet* deletions is presented in Table 1.1.

The first wave of global DNA demethylation occurs in the two gametes following fertilisation (see section 1.2.3 for details). Initially, the maternal genome was shown to

be demethylated passively in a replication-dependent manner through the lack of DNMT1 activity (Rougier et al., 1998; Cardoso and Leonhardt, 1999). The TET3 enzyme, which is highly expressed in the zygote, was thought to be responsible for actively demethylating the paternal genome prior to the onset of cell divisions (Oswald et al., 2000; Gu et al., 2011). The maternal genome is protected from the actions of TET3 via binding of the PGC7 protein to the H3K9me2 histone mark (Nakamura et al., 2012). However, more recent studies indicate that both the maternal and paternal genomes are actively and passively demethylated (Guo et al., 2014; Wang et al., 2014). Lastly, the requirement of TET3 for paternal genome demethylation was recently questioned through demonstration of the formation of 5hmC following the loss of 5mC in the paternal genome (Amouroux et al., 2016). Instead, the authors suggested that an unidentified active DNA demethylation mechanism is responsible for 5mC loss while TET3 plays a role in preventing the gain of aberrant methylation patterns through the

Table 1.1 The summary of the effects of *in vivo Tet* deletions on mammalian development

Knockout type	Phenotype	References
Single *Tet1*	Strain dependent; no developmental defects or partial/complete embryonic lethality	(Dawlaty et al., 2011) (Kang et al., 2015) (Khoueiry et al., 2017)
Single *Tet2*	No developmental defects; myeloid malignancies in adult animals	(Li et al., 2011) (Ito et al., 2019)
Single *Tet3*	Complete embryonic lethality	(Gu et al., 2011)
Double *Tet1* and *Tet2*	Partial embryonic lethality; reduced fertility in surviving animals	(Dawlaty et al., 2013)
Double *Tet1* and *Tet3*	Complete embryonic lethality	(Kang et al., 2015)
Triple *Tet1*, *Tet2* and *Tet3*	Severe gastrulation defects and complete embryonic lethality	(Dai et al., 2016) (Li et al., 2016)

actions of DNMT1 and DNMT3a, which contrary to past findings, were demonstrated to be active in the zygote (Amouroux et al., 2016).

The second wave of DNA demethylation involves imprint erasure in the PGCs. The active removal of cytosine methylation in PGCs occurs following the initial passive demethylation (see section 1.2.3) and focuses on a set of specific genomic regions including some imprinting genes and X-chromosome CGIs (Guibert et al., 2012; Seisenberger et al., 2012; Hackett, et al., 2013). This process was found to be mediated by the TET1 and TET2 enzymes with 5hmC reaching its peak at ED11.5 before being diluted out of the genome by ED13.5 (Hackett, et al., 2013; Vincent et al., 2013; Yamaguchi et al., 2013). At the same time the *Tet3* expression was found to be undetectable (Hackett, et al., 2013). However, while *Tet1* was found to be the highest expressed *Tet* gene during active PGC demethylation, its knockout did not prevent almost complete loss of the 5mC mark by ED13.5 (Hill et al., 2018). Additionally, TET1 was found to play a similar role to TET3 during zygotic demethylation by protecting genomic regions from acquiring aberrant DNA methylation patters (Hill et al., 2018).

These results suggest that the TET enzymes play a role in both active removal of the 5mC mark and protection from unwarranted DNA methylation. Further research is required to determine which of these functions is predominant.

1.4.2 Acquisition of cellular identity and 5-hydroxymethylcytosine - stem cell differentiation and the small intestine

In addition to their critical importance for early mammalian development described in the previous section, the TET enzymes are also known to be important regulators of the stem cell differentiation process. In mESCs, the *Tet1* and *Tet2* transcripts were found to be highly expressed while the *Tet3* mRNA is present at very low levels (Ito et al., 2010; Koh et al., 2011). Upon differentiation to embryonic bodies, the *Tet3* levels rise sharply, *Tet1* levels exhibit a large drop and *Tet2* levels remain largely unchanged following an initial drop (Szwagierczak et al., 2010). In terminally differentiated cells of various tissue types, *Tet2* and *Tet3* levels are high while *Tet1* is barely detectable in most of them (Ito et al., 2010; Szwagierczak et al., 2010).

The distribution of the 5hmC mark in mESCs has been extensively studied and characterised. The total levels of 5hmC are estimated between 0.1% and 0.4% of all cytosines (Le et al., 2011; Münzel et al., 2011; Pfaffeneder et al., 2011; Pfaffeneder et al., 2014). The 5hmC base was found to be enriched in gene-rich areas of the genome, in particular in exons, and around their boundaries, of highly expressed genes and promoters of genes with low expression levels (Xu et al., 2011; Huang et al., 2014). Within the promoter regions, the 5hmC mark was found to specifically be enriched around the transcription start site (TSS) followed by a gradual increase towards the transcription termination site (Xu et al., 2011; Huang et al., 2014). In human ESCs (hESCs), 5hmC was also found to accumulate around the TSS and gene bodies of lowly expressed genes, as well as enhancers and DNA-protein interaction sites (Stroud et al., 2011; Yu et al., 2012; Li et al., 2018).

The experiments with *Tet* knockout mESCs shed further light on their roles in pluripotency maintenance and differentiation (summarised in Table 1.2). A single knockout of the *Tet1* gene resulted in reduced 5hmC levels and a skewing towards the trophectoderm lineage during differentiation (Dawlaty et al., 2011). The *Tet1* knockout also did not alter the pluripotency of the mESCs and it was compatible with embryonic development (Dawlaty et al., 2011). The single deletion of *Tet2* in mESCs resulted in altered expression of genes involved in differentiation to neural progenitor cells which was associated with the hypermethylation and reduced activity of a subset of enhancer regions (Hon et al., 2014). The knockout of *Tet2* resulted in a larger reduction of 5hmC levels (90%) than the *Tet1* knockout (35-44%) (Dawlaty et al., 2011; Hon et al., 2014).

The TET1 and TET2 proteins were found to play distinct roles in mESCs with TET1 mainly responsible for 5hmC deposition in the promoter regions while TET2 drives the accumulation of the 5hmC marks within the genes bodies (Huang et al., 2014). This is thought to be driven by differential targeting of the two enzymes to genomic regions due to the lack of the DNA binding CXXC domain in the structure of the TET2 protein (Huang et al., 2014). Intriguingly, TET1 was found to both activate the expression of genes involved in mESC maintenance through safeguarding of a hypomethylated state and repress genes involved in differentiation through association with the PRC2 complex (Wu et al., 2011). The binding to the PRC2 group of proteins by TET1 was also shown to play a role in maintaining the bivalency of certain promoters in mESCs which is thought to supress their expression (Neri et al., 2013). TET1 also binds to the

SIN3A protein which plays an important role in pluripotency maintenance and their interaction may explain the skewing of differentiation lineage choice in *Tet1* knockout mESCs (Zhu et al., 2018). The single knockout of *Tet3* resulted in skewing of mESC differentiation towards cardiac mesoderm, which was mediated by aberrant activation of the WNT pathway (Li et al., 2016).

Table 1.2 The summary of the effects of *Tet* deletions on mouse embryonic stem cells

Knockout type	Phenotype	References
Single *Tet1*	No effect on mESCs pluripotency and compatible with *in vivo* development	(Dawlaty et al., 2011)
Single *Tet2*	Enhancer hypermethylation; loss of 90% of 5hmC marks and delayed induction of genes involved in differentiation	(Hon et al., 2014)
Single *Tet3*	Aberrant activation of WNT signalling and skewed differentiation towards cardiac mesoderm	(Li et al., 2016)
Double *Tet1* and *Tet2*	Complete loss of 5hmC; no effect on mESCs pluripotency; skewed differentiation towards trophoblast cells	(Dawlaty et al., 2013)
Triple *Tet1*, *Tet2* and *Tet3*	Complete loss of 5hmC in both mESCs and embryonic bodies; severely impaired differentiation with a skew towards cardiac mesoderm, telomere homeostasis and enhancer demethylation	(Dawlaty et al., 2014) (Lu et al., 2014) (Li et al., 2016)

The double knockout of *Tet1* and *Tet2* resulted in complete loss of global 5hmC levels despite elevated expression of *Tet3* (Dawlaty et al., 2013). The authors also demonstrated that the double knockout mESCs are pluripotent and capable of differentiation, albeit defective, into the three germ layers with an aberrant preference towards the trophoblast lineage (Dawlaty et al., 2013). The triple knockout of the *Tets* resulted in complete loss of 5hmC levels but the mESCs retained their normal morphology, pluripotency marker expression and proliferation rate, although the more recent publication by Li *et al* found a significantly reduced proliferation rate in *Tet* triple knockout mESCs (Dawlaty et al., 2014; Lu et al., 2014; Li et al., 2016). However, the triple knockout mESCs displayed a severe skew towards cardiac mesodermal lineage during the differentiation process due to aberrant activation of the WNT pathway (Dawlaty et al., 2014; Li et al., 2016). Additionally, through studying these cells, the TET proteins were demonstrated to demethylate promoters, enhancers and distal regulatory elements in mESCs and their deletion resulted in increased telomere length (Lu et al., 2014). The crucial role of the TET enzymes in regulation of cellular differentiation was recently confirmed by studying triple *TET* knockout hESCs. The TET enzymes were shown to maintain promoter bivalency through hypomethylation and their loss resulted in aberrant differentiation of embryoid bodies (Verma et al., 2018). In particular, the authors demonstrated impaired neuroectoderm differentiation in hESCs due to *PAX6* promoter hypermethylation caused by the triple *TET* knockout (Verma et al., 2018).

The mouse small intestine is an easily accessible model to study adult stem cell differentiation with a well-defined role of the 5mC mark in this process (see section 1.2.4). Cytosine hydroxymethylation, together with the TET proteins, was demonstrated to regulate this process, although the exact role of the latter remains controversial (see section 6.2.1 for more details on the TET enzymes) (Haffner et al., 2011; Kim et al., 2016; Uribe-Lewis et al., 2020). The global 5hmC levels are depleted in the ISCs compared to the differentiated villus progeny in both mouse and human intestine (Uribe-Lewis et al., 2015; Kim et al., 2016; Uribe-Lewis et al., 2020). Kim and co-authors found that 5hmC presence within the gene bodies was positively correlated with gene expression in the two cell types (Kim et al., 2016). In ISCs, 5hmC was associated with genes involved in cellular differentiation and development while in differentiated cells it was enriched in genes responsible for nutrient transport and metabolism (Kim et

al., 2016). Uribe-Lewis and colleagues further confirmed the importance of 5hmC in controlling gene expression during the differentiation process and highlighted its association with both active and bivalent enhancer regions (Uribe-Lewis et al., 2020). Lastly, both research groups found that while the majority of genes show a positive correlation between the change in expression and 5hmC content during differentiation, in a small proportion of genes the presence of 5hmC is associated with transcriptional repression (Kim et al., 2016; Uribe-Lewis et al., 2020).

Taken together, these results indicate the critical role of the TET proteins and the 5hmC mark during the cellular differentiation process and highlight their functions beyond the originally described ones during stem cell pluripotency maintenance and differentiation.

1.4.3 Forever young - the promise of induced pluripotent stem cells and methylation barriers in cellular reprogramming

In 2006 a ground-breaking study published by Takahashi and Yamanaka demonstrated that the use of a cocktail of four transcription factors: OCT4, SOX2, KLF4 and C-MYC allowed a conversion of differentiated cells to a pluripotent, ESC-like state termed induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka, 2006). This discovery was rewarded with a Nobel prize for Physiology and Medicine in 2012 and the so-called "Yamanaka factors" have demonstrated great promise in tackling aging, aiding drug discovery and testing and the field of regenerative medicine (Johnson and Cohen, 2012; Doss and Sachinidis, 2019; Kane and Sinclair, 2019; Haridhasapavalan et al., 2020; Lee et al., 2020). Intriguingly, the TET1 protein can replace OCT4 in the Yamanaka factor cocktail as it demethylates the *Oct4* promoter and enhancer resulting in its reactivation (Gao et al., 2013). The TET1 and TET2 proteins were demonstrated to enhance the efficiency of reprogramming through demethylation of key genomic regions associated with pluripotency such as *Essrb*, *Oct4* and *Nanog* (Costa et al., 2013; Sardina et al., 2018). The critical role of the TET enzymes in the reprogramming process was further highlighted by the fact that the deletion of all three genes in mouse embryonic fibroblasts prevents iPCSs generation (Hu et al., 2014). The authors also suggested that the TETs are critical for the mesenchymal to epithelial transition step of reprogramming rather than the reactivation of the pluripotency genes, which is in direct contradiction to the studies mentioned above (Hu et al., 2014). Vitamin C was shown to

improve the efficiency of mouse and human iPSCs generation (Esteban et al., 2010), although its presence in the media was suggested to prevent TET1 from promoting the reprogramming process (Chen et al., 2013).

1.4.4 The link between cellular differentiation and cancer

The processes of cellular differentiation and tumorigenesis share multiple common characteristics. In clinical pathology, grading of tumour differentiation status has been used for decades as a prognostic marker (Jögi et al., 2012). In general terms, higher degree of differentiation of tumour cells is positively correlated with patient prognosis, while a poorly differentiated tumour that does not resemble the tissue of origin tends to be more malignant and hence difficult to treat (Jögi et al., 2012). One of the four Yamanaka factors involved in generation of iPSCs, C-MYC, is a known oncogene (Takahashi and Yamanaka, 2006). Indeed, teratoma formation, along with tumours arising from cells which differentiated into an unintended tissue types, still represents a major risk for therapies based on pluripotent stem cells (Kumazaki et al., 2013; Itakura et al., 2017; Kimura et al., 2019; Martin et al., 2020). Additionally, the concept of "cancer stem cells" (CSCs) is becoming more widely accepted, with these cells forming a subpopulation within the tumour that is more resistant to conventional treatments such as chemotherapy and radiotherapy which can be responsible for patient relapse (Gupta et al., 2009; Kurrey et al., 2009). These CSCs have now been identified in multiple cancer types including colorectal (Dalerba et al., 2007; Ricci-Vitiani et al., 2007), liver (Ma et al., 2007), breast (Al-Hajj et al., 2003) and brain (Singh et al., 2004) cancers. Lastly, the ISCs are known to give rise to tumours in mice thus providing a further link between the differentiation status and oncogenic transformation potential of the small intestine cells (Barker et al., 2009; Visvader, 2011).

1.5 5hmC and the TETs as important pieces of the tumorigenesis puzzle

1.5.1 The epigenetic hallmarks of cancer cells

In their landmark 2000 review paper published in *Cell*, Douglas Hanahan and Robert Weinberg directed a shift in thinking about the common characteristics of the heterogenous disease of cancer (Hanahan and Weinberg, 2000). The six acquired capabilities, or hallmarks, they described are evasion of apoptosis, self-sufficiency in growth signals, insensitivity to anti-growth signals, sustained angiogenesis, limitless replicative potential and tissue invasion and metastasis (Hanahan and Weinberg, 2000). In an update to the original publication, the authors presented two emerging hallmarks: deregulation of cellular energetics and avoidance of immune destruction as well as two enabling characteristics: genome instability and mutation and tumour-promoting inflammation (Hanahan and Weinberg, 2011). The role of epigenetics and its contribution to establishment of each of the hallmarks in cancer cells is increasingly being recognised (Macaluso et al., 2003; Schnekenburger et al., 2015; Flavahan et al., 2017). It was proposed that dysregulation of epigenetic mechanisms, in particular aberrant DNA methylation, should in fact be considered as a hallmark of cancer development and progression (Stahl et al., 2016; Cheng et al., 2019; Darwiche, 2020). In particular, global loss of 5hmC and promoter hypermethylation of tumour suppressor genes were suggested as epigenetic hallmarks of melanoma (Lian et al., 2012; Mannavola et al., 2020). It is still debatable whether these epigenetic changes should be conceptualised as one of the cancer hallmarks. One could argue that epigenetic changes, as well as DNA mutations, are mechanisms by which the cancer hallmarks are acquired. However, the well-established role of cytosine methylation and accumulating evidence of the importance of the TET enzymes and 5hmC suggest that they should be considered as part of the hallmark framework in the future.

1.5.2 The landscape of 5-hydroxymethylcytosine in the tumour genome

The global loss of 5hmC is a characteristic feature of a number of both solid and haematological cancers (Table 1.3). This has been reported to be in part due to the *TET* mutations present in numerous cancer types (see section 1.5.3) (Pfeifer et al., 2013).

Additionally, the mutations in the isocitrate dehydrogenase genes (*IDH1* and *IDH2*) result in production of an oncometabolite 2-hydroxyglutarate instead of the TET co-factor and Krebs cycle intermediate 2-oxoglutarate (also known as α-ketoglutarate). This results in competitive inhibition of the activity of the TET proteins thus lowering global 5hmC levels (Xu et al., 2011).

Table 1.3 The examples of human cancer types with globally reduced 5hmC levels

Cancer type	References
Bladder	(Peng et al., 2018)
Breast	(Tsai et al., 2015)
Cervical	(Zhang et al., 2016; Bhat et al., 2017)
Colorectal	(Uribe-Lewis et al., 2015; Huang et al., 2016; Hu et al., 2017)
Gastric	(Yang et al., 2013; Park et al., 2015)
Glioma/Glioblastoma	(Johnson et al., 2016; Fernandez et al., 2018)
Head and neck	(Misawa et al., 2019)
Kidney	(Chen et al., 2016)
Leukaemia	(Marçais et al., 2017; Wernig-Zorc et al., 2019)
Liver	(Thomson et al., 2016)
Lung	(Li et al., 2016; Liao et al., 2016)
Lymphoma	(De Souza et al., 2014; Lemonnier et al., 2018)
Melanoma	(Lian et al., 2012)
Myeloma	(Alberge et al., 2019)
Oesophageal	(Murata et al., 2015)
Oral	(Jäwert et al., 2013; Wang et al., 2017)
Ovarian	(Zhang et al., 2015; Tucker et al., 2018)
Pancreatic	(Bhattacharyya et al., 2013)
Prostate	(Feng et al., 2015; Storebjerg et al., 2018)
Testicular	(Munari et al., 2016)
Thyroid	(Tong et al., 2019; Iancu et al., 2020)
Uterine	(Ibrahim et al., 2019)

Tumour hypoxia is also known to lead to reduced hydroxymethylation across the genome and hypermethylation of tumour suppressor gene promoters by reducing the availability of oxygen to the TET enzymes (Thienpont et al., 2016). Interestingly, it is known that the 5hmC loss does not correlate with *TET* and *IDH* mutations in certain cancer types (Jin et al., 2011; Uribe-Lewis et al., 2015). Instead, it was suggested that the reduction in global 5hmC content is due to the increased proliferation rate of cancer cells as the DNMT1 enzyme does not recognise and maintain 5hmC marks and there is a delay in their re-establishment following DNA replication (Valinluck and Sowers, 2007; Otani et al., 2013; Bachman et al., 2014; Uribe-Lewis et al., 2015).

Changes in the 5hmC distribution were also observed in tumours' genomes isolated from patients as compared to matched normal cells. In melanoma the loss of 5hmC was found to occur across all genomic sites, including promoter, intragenic and intergenic regions, compared to nevi cells (Lian et al., 2012). In gastric adenocarcinoma, the biggest loss of 5hmC was observed in exons and CGIs (Liu et al., 2019). Bladder cancer is characterised by a loss of 5hmC content mainly in the promoters and intragenic regions (Peng et al., 2018). In kidney tumours, sites with 5hmC loss were enriched within the gene bodies of genes associated with renal cell carcinoma progression (Chen et al., 2016). Glioblastomas were found to have reduced 5hmC levels within the promoter regions and elevated 5hmC content at enhancer regions (Johnson et al., 2016). The 5hmC enrichment at genomic enhancers was further confirmed in IDH1 wild-type and mutant gliomas (Glowacka et al., 2018). The hyperhydroxymethylation of enhancer elements was also reported in hepatocellular carcinoma patient samples (Hlady et al., 2019). Intriguingly, two studies found 5hmC enrichment in promoters and gene bodies of cancer-associated genes in pancreatic cancer patient-derived xenografts and samples (Bhattacharyya et al., 2013; Bhattacharyya et al., 2017). The increased 5hmC content within these genes was positively correlated with their expression levels (Bhattacharyya et al., 2013; Bhattacharyya et al., 2017).

These results suggest that global 5hmC loss is a hallmark of most cancer types and it predominantly occurs in promoter and intragenic regions during cancer progression. However, in certain cancer types aberrant increase in 5hmC levels is observed at selected genomic loci, most prominently the enhancer elements. It is hence likely that the heterogeneity of pathological genetic and signalling changes between tumour types extends to epigenetic alterations that contribute to their progression and development.

1.5.3 The prevalence of TET and IDH enzyme mutations in human cancers

The fusion of the *TET1* and the HMT *MLL* genes was initially described in acute myeloid leukaemia biopsies harbouring the t(10;11)(q22;23) translocation (Ono et al., 2002; Lorsback et al., 2003). Subsequently, mutations in the *TET* genes were identified in numerous haematological and solid cancers.

In haematological malignancies, loss-of-function *TET2* mutations are the most prevalent mutation type among the three *TET* genes, with *TET1* and *TET3* mutations rarely observed in these cancers (Abdel-Wahab et al., 2009; Huang and Rao, 2014; Lasho et al., 2018). This is despite the latter two genes being expressed in these malignancies, albeit at reduced levels compared to the *TET2* transcript (Scourzic et al., 2015; Lio et al., 2019). The *TET2* gene is mutated particularly often (as high as 70-80% of cases) in angioimmunoblastic (AIML) and peripheral T-cell lymphomas (Lemonnier et al., 2012; Odejide et al., 2014; Palomero et al., 2014), chronic myelomonocytic leukaemia (Kosmider et al., 2009; Tefferi et al., 2009; Haferlach et al., 2014; Xie et al., 2014), blastic plasmacytoid dendritic cell neoplasm (Alayed et al., 2013; Menezes et al., 2014) and acute myeloid leukaemia (AML) (Abdel-Wahab et al., 2009; Shen et al., 2011; Weissmann et al., 2012; Ley et al., 2013).

The mutations in the *TET* genes are rarely observed in solid cancers, up to approximately 10% of cases assessed in most solid tumour types (Scourzic et al., 2015). *TET1* was observed to be mutated in 18.4% of prostate cancer cases (Spans et al., 2016). The analysis of the COSMIC database revealed a lower *TET1* mutation frequency (0.12%) in prostate cancer, likely due to larger sample size compared to the study referenced above (Li et al., 2016; Spans et al., 2016). The latter study also found low *TET1* mutation burden in ovarian (0.57%), lung (0.12%), breast (0.07%) and upper respiratory and digestive system (0.25%) cancers. *TET2* was mutated in 18.2% of neuroendocrine (Pivovarcikova et al., 2019) and 5.7% of clear-cell (Sato et al., 2013) kidney tumours. Its mutations were identified in 10% of endometrial cancers (Chang et al., 2016) with copy number losses reported in the vast majority (73%) of clear cell ovarian cancers (Kim et al., 2018). *TET2* was found to be mutated in 100% of metastatic and 0% of primary breast cancer cases (Nickerson et al., 2013). In gastric cancer, only 2.9% of cases were positive for *TET3* mutations (Mo et al., 2020). *TET3*

was also reported to have a low mutation frequency (<1%) in ovarian cancer (Cao et al., 2019).

The genes encoding the two main IDH enzymes, *IDH1* (cytoplasmic) and *IDH2* (mitochondrial), were described to be mutated in numerous haematological malignancies (Lio et al., 2019). These include AML (15-30%) (Figueroa et al., 2010), AIML (30.4%) (Sakata-Yanagimoto et al., 2014), blast crisis chronic myeloid leukaemia (3.5%) (Soverini et al., 2011) and myelodysplastic syndrome (MDS) (12%) (Patnaik et al., 2012). Intriguingly, *TET2* and *IDH1/2* mutations were found to be mutually exclusive in AML patients, but not in AIML (Figueroa et al., 2010; Sakata-Yanagimoto et al., 2014).

The mutations of *IDH1/2* are more prominent in the more aggressive AML while *TET2* mutations occur with greater frequency in the more benign MDS and myeloproliferative neoplasms (Shih et al., 2012; Xie et al., 2014). This indicates that alterations in *IDH1/2* genes likely have a higher oncogenic potential than *TET2* mutations (Shih et al., 2012; Xie et al., 2014; Inoue et al., 2016). This could be in part due to the inhibition of a wider range of enzymes by *IDH1/2* mutations; indeed, enzymes such as TET1, TET3 (Xu et al., 2011), histone lysine demethylases (Chowdhury et al., 2011; Xu et al., 2011; Janke et al., 2017), transaminases (McBrayer et al., 2018) and prolyl hydroxylases (Xu et al., 2011; Sasaki et al., 2012) were found to be inhibited by 2-hydroxyglutarate.

IDH1/2 mutations were also identified in solid cancers such as glioma, where the mutation burden of these genes in very high (70-90%) (Yan et al., 2009; Dang et al., 2010). Approximately 10% of melanoma cases were found to harbour an *IDH1/2* mutation (Shibata et al., 2011). In chondrosarcomas, alterations in *IDH1/2* genes were reported in between 46.1 and 56% of cases depending on their origin and grade (Amary et al., 2011; Arai et al., 2012). The incidence of *IDH1/2* mutations is lower in malignancies such as thyroid (7.9%) (Murugan et al., 2010) and lung (0.5%) (Rodriguez et al., 2020) cancers.

These results highlight the considerable variation in *TET* and *IDH* mutation prevalence in human cancers. Furthermore, they point to the fact that these genetic alterations on their own are not responsible for all the hallmark global 5hmC loss observed in most cancer types.

1.5.4 The TET enzymes in cancer progression and metastasis

Since the discovery of TET1 and its ability to convert 5mC to 5hmC in 2009, the levels and activity of the TET enzymes have been studied in numerous cancer types (Tahiliani et al., 2009; Ko et al., 2010; Lian et al., 2012; Carella et al., 2020; Yu et al., 2020). In melanoma patients, the expression of all three *TET* genes was found to be reduced in the melanoma lesions compared to benign nevi, with *TET2* levels showing the greatest reduction (Lian et al., 2012). In glioblastoma patients, the expression of *TET3* was shown to be reduced (Carella et al., 2020) while *TET1* and *TET2* expression was lower compared to normal tissue in papillary thyroid carcinoma sufferers (Yu et al., 2020). Similarly, the expression of *TET1* was found to be lower in pancreatic tumour tissue compared to normal tissue (Wu et al., 2019). There is evidence, however, that in certain cancer types, such as lung cancer and leukaemia, TET1 can also play an oncogenic role (Huang et al., 2013; Filipczak et al., 2019) Additionally, *TET3* expression was shown to be elevated and predict poor patient outcome in renal cell carcinoma (Chen et al., 2017) and ovarian cancer (Cao et al., 2019). These results indicate that while the TET enzymes broadly have a tumour suppressive function, this pattern may not be true in all cancer types and their role should be investigated in each tumour type separately.

The mostly tumour suppressive function of the TET enzymes is further supported by studies assessing the effects of altering the levels of these proteins both *in vitro* and *in vivo* in multiple cancer types. *TET1* transcript overexpression was found to reduce cell growth and promote apoptosis in an osteosarcoma cell line (Teng et al., 2019). Similarly, overexpression of TET1 led to reduced cell proliferation in hepatocellular carcinoma cell lines (Lin et al., 2015). Overexpression of the TET2 enzyme in a melanoma cell line resulted in reduced invasion *in vitro* and reduced tumour growth *in vivo* (Lian et al., 2012). Furthermore, overexpression of the TET2 protein resulted in reduced growth in gastric cancer (Deng et al., 2016) and glioblastoma (García et al., 2018) cell lines. Similarly, the overexpression of the TET3 enzyme was shown to have tumour suppressive properties in ovarian cancer (Ye et al., 2016) and glioblastoma (Carella et al., 2020) cell lines.

The genetic knockouts of the TET enzymes provide further insights into their role in different cancer types. The deletion of *Tet1* gene in mice was shown to predispose the animals to development of B cell lymphoma (Cimmino et al., 2015). The *TET1*

knockout in a pancreatic cancer cell line resulted in increased cell proliferation and migration, thus further supporting the tumour suppressive properties of TET1 (Wu et al., 2019). However, a deletion of *TET1* in a triple negative breast adenocarcinoma cell line led to significant reduction in cell proliferation and migration, indicating a potential oncogenic role of TET1 in some cancer types (Good et al., 2018). A genetic knockout of *Tet2* in a mouse model resulted in development of myeloid lineage cancers in line with the evidence from cancer patients supporting the tumour suppressive role of TET2 in haematopoietic cancers (Li et al., 2011). The knockout of *Tet2* in a murine melanoma cell line did not result in significant changes in tumour growth *in vitro* or *in vivo* but it led to resistance to PD-L1-based immunotherapy (Xu et al., 2019). Additionally, genetic knockout of *TET2* resulted in increased parathyroid cancer cell growth and metastasis *in vitro* (Barazeghi et al., 2017). Mice with a double *Tet2* and *Tet3* knockout were shown to develop aggressive myeloid leukaemia which was not observed in a single knockout in ether of the genes, indicating synergistic tumour suppressive effects of TET2 and TET3 in mice (An et al., 2015). This result was further supported by genetic knockouts of three out of four *Tet2* and *Tet3* alleles which also resulted in development of acute myeloid leukaemia (Shrestha et al., 2020). Additionally, *TET3* knockdown in tumour-initiating cell subpopulations of two breast cancer cell lines reduced tumour growth *in vitro* and *in vivo* (Wu et al., 2015). The shRNA-mediated *TET3* knockdown in a glioblastoma cell line, on the other hand, was found to increase tumour growth both *in vivo* and *in vitro* (Cui et al., 2016).

Taken together, the findings described in the above sections present a highly complex story of the role of the TET proteins and 5hmC in cancer progression and metastasis. Despite the progress made in the field, numerous questions remain. These include, but are not limited to, the following. Is the global loss of 5hmC levels observed in most cancer types mostly due to their increased proliferation rate? If not, what are the exact mechanisms responsible for global hypohydroxymethylation in cancers without a large burden of *TET* and *IDH* mutations? Are reduced 5hmC levels a by-product of cellular transformation or do they drive cancer progression? How do the metabolic changes characteristic of cancer cells affect and regulate the catalytic and non-catalytic functions of the TET proteins? What do the dual context-dependent roles, both as tumour suppressors and oncogenes, of the TET enzymes reported so far mean for their viability as drug targets? Do 5fC and 5ca also have potential roles in cancer progression?

1.6 Colorectal cancer as a disease of modernity and its epigenetics

1.6.1 Colorectal cancer in the contemporary world

Colorectal cancer (CRC) is described as "disease of modernity" due to its association with improved standard of living and the fact that it mainly occurs in developed countries (reviewed in Brody, 2015; Holmes, 2015). In the USA alone, the projections estimate 147,950 cases and 53,200 death will be recorded in 2020 (Siegel et al., 2020). In 2018, approximately 1.8 million new cases were diagnosed worldwide and 881,000 deaths were attributed to the disease which account for just over 10% of all cancer cases and deaths that year (Bray et al., 2018). The incidence of CRC in developed countries such as the USA, UK, Australia, Denmark, New Zealand and Canada is either stabilising or declining, although the opposite trend is observed in adults below 50 years of age in these nations (Araghi et al., 2019; Siegel et al., 2020). The decline of CRC incidence in those countries is mainly attributed to improved uptake and availability of screening programmes, higher standard of healthcare as well as dietary and lifestyle changes (Ouakrim et al., 2015; Mármol et al., 2017). Despite this, the yearly global burden of cases and deaths due to CRC is predicted to rise to 2.2 million and 1.1 million, respectively, in 2030 (Arnold et al., 2017) and 2.5 million cases by 2035 (Dekker et al., 2019; Xie et al., 2020). These models stress the importance of improvements in prevention, access to healthcare and screening programmes and development of novel therapeutic agents to combat the disease.

1.6.2 Colorectal cancer as a model for stepwise tumour progression and its metastasis

The aetiology of CRC has clear and well-studied familial and genetic components: approximately 70% of cases are sporadic, 25% are of familial origin without a known germline mutation and 5% are due to clearly defined, inherited genetic changes associated with hereditary syndromes (Mármol et al., 2017; Jansen et al., 2020; Schubert et al., 2020). The latter include the Lynch Syndrome, Familial Adenomatous Polyposis (FAP) and a group of Hamartomatous Polyposis Syndromes (Stoffel and Kastrinos, 2014). Recently, a number of previously unidentified candidate genes have

been suggested to contribute to the familial cases of CRC, including *POLD1*, *NOTCH2*, *RAB25*, *GALTN12* and *RPS20* (Jansen et al., 2020; Schubert et al., 2020).

In the sporadic cases, the development of CRC is proposed to arise through a defined series of genetic mutations which result in progression from a normal colonic epithelium through adenoma to an invasive carcinoma. This hypothesis was first put forward in 1990 and centred around *KRAS* and *TP53* mutations (Fearon and Vogelstein, 1990). It is now generally accepted that the initial mutation in the progressive sequence is observed in the adenomatous polyposis coli (*APC*) gene (Vogelstein et al., 2013). *APC* is a known tumour suppressor gene which plays a role in ß-catenin degradation and its mutation is responsible for the FAP inherited syndrome (Leoz et al., 2015). Its initial mutation results in an increase in cell proliferation rate and formation of small adenomas which progress to larger ones through mutations in *KRAS* or *BRAF* oncogenes (Huang et al., 2018). The mutations in *TP53*, *DCC*, *SMAD4* and *PIK3CA* genes occur at a later stage in the malignant progression sequence, leading to development of carcinomas (Fearon and Vogelstein, 1990; Vogelstein et al., 2013; Huang et al., 2018). The *APC*, *KRAS* and *TP53* genes are mutated in 81%, 43% and 60% of sporadic CRC cases which highlights their importance in CRC progression (Lichtenstern et al., 2020).

The most common site of CRC metastasis is the liver which occurs due to the direct drainage of blood containing nutrients from the intestines through the hepatic portal vein (Zarour et al., 2017). The second common site of CRC metastases are the lungs, with the peritoneum, the brain, bones, adrenal glands and the spleen representing less commonly observed targets for CRC metastases (Vatandoust et al., 2015). Approximately 15-20% and over 50% of CRC patients develop lung and liver metastases, respectively, at some stage of their life (Misiakos et al., 2011; Valderrama-Treviño et al., 2017). Only 20-25% of CRC patients presenting with liver metastases qualify for their resection (Vatandoust et al., 2015; Engstrand et al., 2018). Following hepatic resection, the 5-year survival is estimated at approximately 45% (Wang et al., 2017) The prognosis is much worse for patients with non-resectable liver metastases who have a 5-year survival of below 10% (Gorgen et al., 2018; Dueland et al., 2020).

The genes implicated in CRC progression and described in the previous paragraph are also closely linked to the metastatic process - a significant association was found

between *KRAS*, *SMAD4* and *TP53* mutations and CRC metastasis (Huang et al., 2018). The exact details of the role of these genes in the metastatic cascade remains, however, unclear. Interestingly, the role of CRC CSCs in the metastatic process is gaining increasing support (Zhou et al., 2018). The CRC CSCs subpopulations have previously been reported to express cell surface markers such as CD26, CD44, CD133, CD166, ALDH1, EpCAM and LGR5, which allow for their isolation and study (Dalerba et al., 2007; Pang et al., 2010; De Sousa E Melo et al., 2017; Zhou et al., 2018). The CD26-positive cells were found present in liver metastases of CRC patients and were capable of forming metastases in mice (Pang et al., 2010). Similarly, CSCs expressing variant 6 of CD44 membrane protein were the drivers of CRC metastasis (Todaro et al., 2014). The LGR5-positive CSCs were shown to be necessary for CRC metastasis formation (De Sousa E Melo et al., 2017). A recent study further demonstrated that the CRC metastasis is driven by LGR5-negative cells, but the growth at the secondary site requires them to convert to the CSC LGR5-positive phenotype (Fumagalli et al., 2020). This indicates that the presence of CSCs and a degree of plasticity in the differentiation phenotype of CRC cells is required for a successful distant metastasis.

1.6.3 The epigenetics of colorectal cancer

The role of cytosine methylation in CRC development and progression has been extensively studied to date (Lam et al., 2016; Vymetalkova et al., 2019; Cervena et al., 2020). Both hypo- and hypermethylation is observed in CRC. The potential role of hypomethylation of CRC was demonstrated as early as 1983 through description of hypomethylation of the *RAS* oncogenes which can lead to their aberrant activation (Feinberg and Vogelstein, 1983b). Hypomethylation of the C-MYC oncogene was also reported in CRC (Luo et al., 2010). Another example involves hypomethylation of the LINE-1 transposable elements in CRC, which comprise just under 20% of the genome and whose hypomethylation is thought to result in chromosomal rearrangements and genomic instability (Ogino et al., 2008; Jung et al., 2020). Conversely, hypermethylation of numerous tumour suppressor genes including *APC*, *CDKN2A*, *MGMT* and *MLH1* was reported in CRC (Vymetalkova et al., 2019; Jung et al., 2020). This gain in cytosine methylation in tumour suppressor gene promoters is associated with the CIMP phenotype, briefly described in section 1.2.6. Inhibitors of DNMT

enzymes are being currently evaluated in early clinical trials for use in CRC patients with higher, aberrant DNA methylation levels (Cervena et al., 2020). Beyond cytosine methylation, the function of TET enzymes and 5hmC in CRC is also increasingly better understood (Chapman et al., 2015; Rawłuszko-Wieczorek et al., 2015; Tian et al., 2017; Gao et al., 2019). In particular, the use of 5hmC as a CRC biomarker is an exciting new avenue in CRC diagnosis. A number of assays have been demonstrated to be capable of distinguishing heathy and cancerous colonic tissue based on global 5hmC levels and distribution (Gilat et al., 2017; Gao et al., 2019; Margalit et al., 2020). Intriguingly, the same could be achieved using cell-free DNA which represents a much less invasive, and hence preferable, form of diagnosis (Li et al., 2017; Gao et al., 2019; Berger et al., 2020). However further validations of the specificity and sensitivity of these methods will be required before they are translated into clinical use. For a detailed discussion of the role of TET proteins in CRC, see section 5.2.

Other components of the epigenetic machinery, such as HDACs, histone demethylases (HDMs), miRNAs and long non-coding RNAs (lncRNAs) have also been demonstrated to play a role in CRC progression (reviewed by Jung et al., 2020). A number of HDAC genes, including *HDAC1*, *HDAC2* and *HDAC8* were found to have elevated expression in CRC and contribute to silencing of tumour suppressor genes (Mariadason, 2008; Chou et al., 2011). HDAC inhibitors have shown promising results in clinical trials with CRC patients and are being investigated further (Azad et al., 2018; Advani and Kopetz, 2019). The *LSD1* HDM was demonstrated to be overexpressed in CRC patients and its inhibition reduces CRC cell line migration *in vitro* (Li et al., 2019; Peng et al., 2020). The miRNAs, such as miRNA-21 and miRNA-92a which are upregulated in CRC, have mainly shown promise as diagnostic biomarkers, although their specificity remains controversial (Ren et al., 2015; Fu et al., 2018; Choi et al., 2019; Igder et al., 2019; Sabry et al., 2019). The antisense oligonucleotides targeting these miRNAs, despite showing early promise, are unlikely to be used in clinical setting anytime soon (Ding et al., 2018; Jung et al., 2020). The lncRNAs can function both as tumour suppressors, such as the *GAS5* lncRNA (Liu et al., 2019; Ni et al., 2019), and oncogenes such as the *CCAT1* lncRNA (Abedini et al., 2019; Chen et al., 2019). While certain lncRNAs, such as *CCAT1* and *CCAT2* show early promise as CRC biomarkers, the current evidence of their sensitivity and specificity is too sparse to justify clinical implementation in the near future (Zhao et al., 2015; Ozawa et al., 2017).

In recent years the research of epigenetic influences on CRC development and progression provided an understanding of this process that goes beyond the genetic model of colon epithelium-to-adenoma-to-carcinoma sequence. Further research in this area should provide exciting discoveries in both detection and treatment of CRC in the future.

1.7 The main hypothesis and key aims of the project

The main hypothesis:

The TET enzymes and 5hmC contribute to the process of colorectal cancer metastasis to the liver.

Key aims of the project:

1. To develop a cost-effective sensitive method of detection for global 5mC and 5hmC levels using the mass spectrometry facilities available at the University of Bath.
2. To characterise the *in vitro* properties of the two colorectal cancer cell lines, SW480 and SW620, selected as a metastasis model and the effects of compound-based 5hmC levels modulation on the *in vitro* migration and proliferation rates of these cells.
3. To examine the effects of CRISPR/Cas9-induced mutations in genes encoding all three TET enzymes on *in vitro* proliferation, colony formation and migration of the SW480 and SW620 colorectal cancer cell lines.
4. To examine the changes in 5hmC and *TET* levels during adult and embryonic stem cell differentiation of the small intestine and the liver to further the understanding of their role in untransformed tissue and to complement the results obtained with the *in vitro* colorectal cancer model.

Chapter 2

Materials and methods

2.1 Materials

The reagents and chemicals used in this project were purchased from Sigma-Aldrich, Thermo Fisher Scientific or VWR unless otherwise specified. The primers were purchased from Eurofins Genomics and Primerdesign.

2.1.1 Cell lines

The HCT116 cell line was obtained from American Type Culture Collection.

The SW480 and SW620 cell lines were obtained from Professor Ann Williams (School of Cellular and Molecular Medicine, University of Bristol).

The SW480 2I7, SW480 6B12 and SW620 2H13 TET triple knockout cell lines were obtained from Horizon Discovery.

The H9 human embryonic stem cell line was obtained from Professor David Tosh (Department of Biology and Biochemistry, University of Bath)

2.1.2 Murine samples

The mouse brain tissue was obtained from Dr Kim Moorwood (Department of Biology and Biochemistry, University of Bath).

The FACS-sorted CD24a_Mid (stem progenitor cells) and CD24a_Neg (differentiated progeny cells) cell aliquots isolated from mouse small intestines were obtained from Dr Anna Nicholson (Cancer Research UK Institute, University of Cambridge).

2.2 Cell culture

All cell lines were cultured in a 37°C incubator with CO_2 levels maintained at 5% in a humidified environment.

2.2.1 Cell lines passaging

The HCT116 cells were cultured in the Roswell Park Memorial Institute-1640 (Sigma-Aldrich) supplemented with 5% (v/v) Fetal Bovine Serum (FBS) (Thermo Fisher Scientific) and 1% (v/v) Penicillin-Streptomycin solution (Gibco).

The SW480, SW620, SW480 2I7, SW480 6B12 and SW620 2H13 cell lines were grown in Dulbecco's Modified Eagle Medium containing high glucose concentration of 4500 mg/L (Sigma-Aldrich) supplemented with 5% (v/v) FBS and 1% (v/v) Penicillin-Streptomycin solution.

All cell lines were passaged every three or four days upon reaching 70-85% confluency. This was performed by aspirating the growth media from the culture dish, washing the cell monolayer once with pre-warmed 1X phosphate buffered saline (PBS) (Gibco) and incubating the cells with 0.05% trypsin solution (Gibco) for 8-12 minutes at 37°C. The trypsin solution was then inactivated by adding an appropriate amount of growth media and the desired proportion of the cells were added to the new cell culture dish.

The H9 hESC line was cultured in feeder-free conditions as separate colonies in 6-well cell culture plates. The H9 cells were maintained in either mTeSR1 or mTeSR+ growth media (STEMCELL Technologies) with media changed either daily or every two days, respectively. The extracellular matrix coating for stem cell growth was prepared using human recombinant Laminin-521 (BioLamina) according to the manufacturer's instructions. Briefly, 1 mL/well of the 5 µg/mL Laminin-521 solution prepared in DPBS (Sigma-Aldrich) was added to each plate followed by a 2 hour incubation at 37°C or storage at 4°C overnight before use.

The colonies formed by the H9 cells were passaged upon reaching 70-80% confluency approximately every 3-4 days. The colonies were dissociated using the ReLeSR reagent (STEMCELL Technologies) by washing each well with DPBS and adding 0.5 mL/well of ReLeSR reagent for approximately 30 seconds. Following this short incubation, the

dissociation reagent was removed from each well and the plates were incubated for 5 minutes at 37°C. Following that, 1.5 mL/well of either mTeSR media was added and the sides of the plates were tapped to aid the dissociation of the colonies. The resulting cell suspension was split approximately 1:6 into wells of a new Laminin-521 coated cell culture plate for propagation.

The passaging and maintenance of the H9 hESC cell line was performed by Dr Stephen Weston and Dr Zoë Burke.

2.2.2 Cell lines freezing and thawing

The cell lines used in this project were cryopreserved by detaching the cells from the cell culture dish as described in section 2.2.1. The cell suspension was centrifuged at 400 x g for 5 minutes at RT. The resulting supernatant was aspirated off and the cell pellets were resuspended in freezing media consisting of 50% (v/v) cell culture medium, 40% (v/v) FBS and 10% (v/v) dimethyl sulphoxide (DMSO) (Sigma-Aldrich). The cryovials containing the cells were then tightly wrapped in thick tissue paper and kept at -80°C for 24 hours before transferring them to the liquid nitrogen tank for long term storage.

The cells were thawed by warming the cryovial up in a 37°C water bath until half of its contents reached a liquid state. And equal volume of pre-warmed growth media was added to the cell suspension and gently mixed. The whole mixture was then transferred to a cell culture dish containing 5-10 mL of pre-warmed growth medium. The cells were then allowed to attach to the surface of the culture dish for at least 8 hours, following which the media was changed to remove the trace quantities of DMSO.

2.2.3 Cell treatments with 5hmC levels modifying compounds and assessment of cell morphology

The SW480 and SW620 cell lines were treated with the following compounds known to modify 5hmC levels: vitamin C (L-ascorbic acid) (Sigma-Aldrich), vitamin C and catalase (Sigma), dimethyl 2-oxoglutarate (Sigma-Aldrich) and nickel (II) chloride (Sigma-Aldrich). The doses of the compounds used in these treatments are listed in

Table 2.1. Cell images were taken using the EVOS FL microscope (Invitrogen) at the following 24 hour intervals after the start of the treatment: 0 hours, 24 hours and 48 hours.

Table 2.1 The doses of the 5hmC modifying compounds used in the *in vitro* assays

Compound	Manufacturer	Catalogue number	Concentrations used
L-Ascorbic acid	Sigma-Aldrich	A4403	0.1 mM, 0.5 mM, 1 mM
L-Ascorbic acid and 100 µM Catalase from bovine liver	Sigma-Aldrich	C1345	0.1 mM, 0.5 mM, 1 mM
Dimethyl 2-oxoglutarate	Sigma-Aldrich	349631	1 mM, 2.5 mM, 5 mM
Nickel (II) chloride	Sigma-Aldrich	654507	50 µM, 100 µM, 200 µM

2.2.4 Cell counting

The cell lines were detached from the cell culture dish and dissociated into a single cell suspension using trypsin solution followed by its inactivation with the growth media. The cells were counted using a Neubauer counting chamber following a 1:2 dilution with a trypan blue solution (Sigma-Aldrich). The cell count was obtained by counting the cells in each of the four corner squares out of the nine large squares present in the haemocytometer ruled chamber area. The following equation was used to derive the number of cells per 1 mL:

Total cell count in the four corner squares ÷ 4 × 2 × 10,000

2.2.5 Cell size measurements

The cell images at 20X magnification were taken using the EVOS FL microscope (Invitrogen) 48 hours after seeding. Between 100 and 200 cells per condition and cell line were randomly selected. The size measurement was taken of each cell by measuring the longest possible diameter of each cell using the ImageJ software.

2.2.6 Cell viability assessment - the MTT assay

The cell viability of the cell lines treated with the 5hmC levels modifying compounds and the TET triple knockout cell lines was assessed by the 3-[4,5-dimethylthiazole-2-yl]-2,5-diphenyltetrazolium bromide (MTT) assay.

The cells were seeded in a 96-well cell culture plate at the density of 10,000 cells/ well for all assays. In the case of treatments, the media was changed 24 hours after seeding and this was deemed the "0 hour" time point. The cell viability was assessed 48 hours following the start of the treatment or seeding. In the vitamin C and vitamin C in the presence of catalase experiments the media was changed every 24 hours in all wells.

The media was removed from each well and 1 mg/mL of Thiazolyl Blue tetrazolium bromide (Alfa Aesar) dissolved in 1X PBS was added to all wells. The plates were incubated for 30 minutes at 37°C. The contents of the wells were decanted and isopropanol was added to dissolve the blue formazan crystals. Control wells containing isopropanol-only were included in all experiments to assess the background absorbance. The plates were gently shaken until all crystals were dissolved. The absorbance (A560 nm) was measured using a plate reader (Turner Biosystems).

2.2.7 Cellular transfection

The transfection of of hESCs and cancer cell lines with plasmid constructs was carried out using the FuGENE HD transfection reagent (Promega Corporation) according to manufacturer's instructions. The transfection conditions were obtained from the online FuGENE HD Protocol Database. The details of plasmids used for cellular transfections are shown in Table 2.2.

Table 2.2 The details of the plasmid vectors used for cellular transfections

Plasmid name	Vector type	Vector backbone	Insert (5'→3')	Catalogue number	Source
FH-TET1-pEF	Over-expression	pEF1a	Human *TET1* gene	Addgene #49792	A gift from Anjana Rao (Tahiliani et al., 2009)
FH-Tet2-pEF	Over-expression	pEF1a	Mouse *Tet2* gene	Addgene #41710	A gift from Anjana Rao (Ko et al., 2010)
FH-TET3-pEF	Over-expression	pEF1a	Human *TET3* gene	Addgene #49446	A gift from Anjana Rao (Ko et al., 2013)
Luciferase shRNA	Knockdown (shRNA)	pGSH1-GFP	Forward: GATCCGATTATGTCCGGTTATGTAGAAGCTTGTACATAACCGGACATAATCTTTTTTGGAAGC Reverse: GGCCGCTTCCAAAAAAGATTATGTCCGGTTATGTACAAGCTTCTACATAACCGGACATAATCG	-	A gift from Keith Vance (University of Bath)
TET1 shRNA 1	Knockdown (shRNA)	pGSH1-GFP	Forward: GATCCCAAGTCTCTTATCGTTAATGAAGCTTGATTAACGATAAGAGACTTGTTTTTTGGAAGC Reverse: GGCCGCTTCCAAAAAACAAGTCTCTTATCGTTAATCAAGCTTCATTAACGATAAGAGACTTGG	-	Made and designed by the author
TET1 shRNA 2	Knockdown (shRNA)	pGSH1-GFP	Forward: GATCCCCTCCAGTCTTAATAAGGTTAGAAGCTTGTAACCTTATTAAGACTGGAGGTTTTTTGGAAGC Reverse: GGCCGCTTCCAAAAAACCTCCAGTCTTAATAAGGTTACAAGCTTCTAACCTTATTAAGACTGGAGGG	-	Made by the author Insert designed by (Deng et al., 2017)

2.3 *In vitro* cancer hallmark assays

2.3.1 Cell proliferation assays

The cells were seeded in a 12-well cell culture plate at the density of 50,000 cells/ well for the assessment of the effects of the 5hmC levels modifying compounds on cell proliferation and 100,000 cells/well for all other assays. In the case of treatments, the media was changed 24 hours after seeding and this was deemed the "0 hour" time point for the cell counting purposes. The cells were counted at following 24 hour intervals after the start of the treatment or seeding: 24 hours, 48 hours, 72 hours and 96 hours. In the vitamin C and vitamin C in the presence of catalase experiments the media was changed every 24 hours in all wells.

2.3.2 Cell migration assays

The cells were seeded at a sufficiently high number to allow them to reach confluency within 48 hours in a 6 cm cell culture dish. Upon reaching confluency, the media in the 6 cm dishes was changed to one containing 4 μg/mL Mitomycin C (Sigma-Aldrich) in order to arrest cell proliferation. Following a 2 hour incubation at 37°C, the cells were scratched with a P200 pipette tip as demonstrated in Figure 2.1. Each plate was washed three times with the growth media to remove the dead cells and traces of Mitomycin C.

The 0 hour images were then taken using the EVOS FL microscope as described in Figure 2.1. Pictures of the same areas of the plate were also taken 24 hours and 48 hours after the scratch was performed. In the vitamin C and vitamin C in the presence of catalase experiments the media was changed every 24 hours in all cell culture dishes. The size of the gap at each time point was measured using the ImageJ software for all images by taking an average of three measurements for each image taken, yielding a total of fifteen measurements for each timepoint.

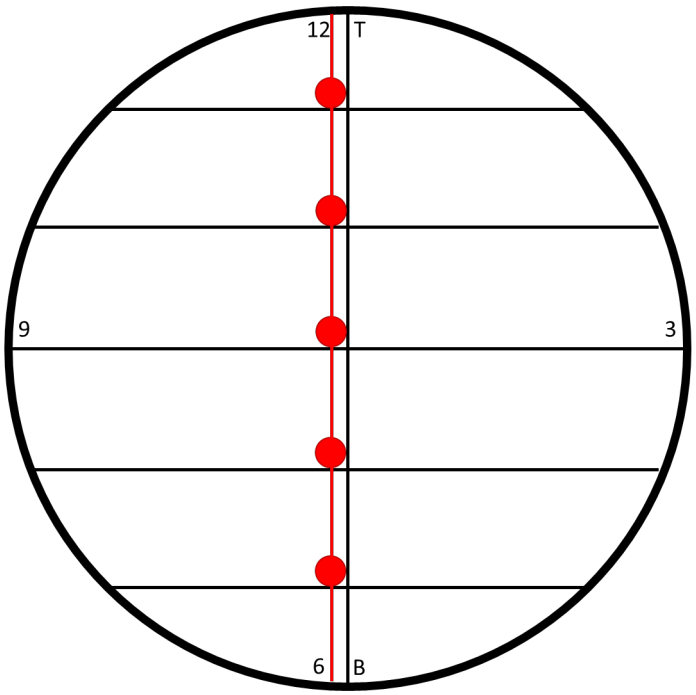


Figure 2.1 The schematic representation of the migration assay set up in a 6 cm cell culture dish
The letters "T" and "B" denote top and bottom of the plate respectively. The red line passing through the red dots represents the scratch made with the pipette tip while the dots themselves mark the areas where images of the gap between the two sides of the scratch were taken. In the schematic the 6 cm dish is depicted upside down.

2.3.3 Colony formation assays

The cells were seeded at the density of 200 cells/well in a 6-well cell culture plate for all assays. After the period of 24 hours, media was changed to one containing the correct concentration of the compound under investigation for the 5hmC levels modifying compounds. In the vitamin C and vitamin C in the presence of catalase experiments the media was changed every 24 hours in all cell culture dishes. For all other assays, the media was changed every 48-72 hours. The growth of the colonies was closely monitored to avoid separate colonies merging together while allowing sufficient time for the colonies to reach quantifiable size.

The plates were fixed by carefully removing all of the growth media from each well and adding the fixing solution (1% v/v methanol and 1% w/v formaldehyde in 1X PBS) for 30 minutes at RT. The colonies were then stained with 0.01% (w/v) crystal violet

solution for 1 hour at RT. Images of each plate were taken using the ChemiDoc-It^2 UV Imager (Ultra-Violet Products). The GIMP software was used to remove any background staining from the images and the ImageJ software was utilized for the purpose of both counting colonies and measuring their surface area in square pixels through the "Analyse particles" macro.

2.4 Epigenetic modulation of human embryonic stem cell differentiation

2.4.1 Human embryonic stem cell differentiation to hepatocyte-like cells

The H9 hESCs were differentiated into hepatocyte-like cells (HLCs) using a modified version of previously published protocols (Hay et al., 2008; Cameron et al., 2014; Wang et al., 2017). Colonies formed by the H9 cells were washed with DPBS and dissociated for 5-10 minutes at 37°C with the Accutase solution (STEMCELL Technologies) added at 0.5 mL/well. An equal volume of the mTeSR+ medium was added to the resulting cell suspension which was then centrifuged for 3 minutes at 1000 rpm. The supernatant was carefully aspirated off and the cell pellet was resuspended in the mTeSR+ medium with the 1X RevitaCell supplement added. The cells were counted as described in section 2.2.4. and seeded at the density of 7 x 10^4 cells/cm^2 in pre-made Laminin-521 coated 6-well cell culture plates.

Approximately 24 hours after seeding, the first stage of the protocol was initiated - namely differentiation into definitive endoderm (DE) cells. The mTeSR+ medium was changed to RPMI-1640 supplemented with the 1X B27 supplement without vitamin A (Thermo Fisher Scientific), 2 μM of the 1m glycogen synthase kinase 3 (GSK3) inhibitor and 100 ng/mL of Activin A (PeproTech). Both 24 and 48 hours after that, the media was changed to RPMI-1640 supplemented with the 1X B27 supplement without vitamin A and 100 ng/mL of Activin A.

Upon reaching the DE stage, the cells were further differentiated into hepatoblast (HB) cells by replacing the media with KnockOut-DMEM (Gibco) supplemented with 20% (v/v) KnockOut Serum Replacement (Gibco), 1% (v/v) Penicillin/Streptomycin

solution, 1 % (v/v) DMSO, 1X MEM Non-Essential Amino Acids Solution (Gibco), 0.5X GlutaMAX supplement (Gibco) and 0.1 mM β-mercaptoethanol solution (Gibco). Subsequently, the same media was used for media changes every 24 hours for the next three days.

The final stage of the protocol involved differentiation of HB cells to HLCs. The media was changed to HepatoZYME-SFM (Gibco) serum-free media supplemented with 1% (v/v) Penicillin/Streptomycin, 1X GlutaMAX supplement, 0.1 mM β-mercaptoethanol solution, 10 ng/mL Recombinant Human Hepatocyte Growth Factor (PeproTech), 20 ng/mL Recombinant Human Oncostatin M (PeproTech), 10 µM hydrocortisone 21-hemisuccinate sodium salt (Sigma-Aldrich), 10 µg/mL human insulin (Sigma-Aldrich), 6.7 ng/mL sodium selenite (Sigma-Aldrich) and 5.5 µg/mL human holo-transferrin (Sigma-Aldrich). Subsequently, the same media was changed every 24 hours for the next 9 days. On the last day (day 18) the cells were pelleted as described in section 2.2.2. and snap frozen at -80°C prior to DNA and RNA extraction.

The differentiation of the H9 hESC cell line to HLCs was performed by Dr Stephen Weston.

2.4.2 Human embryonic stem cell transfection

The H9 hESCs were seeded as described in section 2.4.1. and allowed to proliferate for 24 hours. Following the incubation period, the cells were transfected with 3µg/well of the appropriate shRNA vector using the FuGENE HD transfection reagent according to manufacturer's instructions. A transfection reagent-only control was included in all experiments. After 24 hours, the transfection efficiency was assessed using the EVOS FL microscope and the differentiation to DE stage was initiated and carried out as described in section 2.4.1.

2.4.3 Antibody staining of the definitive endoderm markers

Upon reaching the DE stage, the media was aspirated from wells of the 6-well cell culture dishes and the cells were detached using the Accutase solution added at 1mL/well for 5 minutes at 37°C. Following the incubation, 5 mL/well of 2% (v/v) FBS

in DPBS was added and the cell suspension was centrifuged at 1000 rpm for 3 minutes. The cells were then resuspended in 2% (v/v) FBS in DPBS and split into four tubes for antibody staining.

The anti-c-Kit (anti-CD117) and anti-CXCR4 (anti-CD184) antibodies were added to the appropriate tubes and incubated at 4°C for 30 minutes. The cells were span down at 4°C and 1000 rpm for 3 minutes, washed once in 300 µL of 2% (v/v) FBS in DPBS and centrifuged in the same way again. The cell pellet was resuspended in 500 µL of 2% (v/v) FBS in DPBS and the staining results were analysed using fluorescence-activated cell sorting (FACS). The details of the anti-c-Kit and anti CXCR4 antibodies are listed in Table 2.3.

Table 2.3 The details of the antibodies used for H9 hESCs staining

Antibody name	Catalogue number	Manufacturer	Species of origin	Dilution
c-Kit Monoclonal Antibody (104D2), APC	CD11705	Invitrogen	Mouse	1:100
CXCR4 Monoclonal Antibody (12G5), PE	MHCXCR404	Invitrogen	Mouse	1:100

2.4.4 Fluorescence-activated cell sorting

The FACS analysis was performed on the BD FACSAria III instrument (Becton, Dickinson and Company) according to manufacturer's instructions with the assistance of Dr Michael Zachariadis. The data was interpreted using the BD FACSDiva software (Becton, Dickinson and Company).

2.5 Assessment of global 5mC and 5hmC levels in genomic DNA

2.5.1 DNA extraction

The DNA was extracted from cell lines and tissues using the following protocol unless otherwise specified. The harvested cells or shredded tissue were incubated in 1X PCR buffer (10 mM Tris, 50 mM KCl, 2 mM $MgCl_2$ and pH 8.3) diluted from a 10X stock solution and supplemented with Proteinase K (Sigma-Aldrich) at 65°C for 1 hour. This was followed by heat inactivation of the Proteinase K enzyme for 15 minutes at 95°C. An equal volume of Phenol:Chloroform:Isoamyl alcohol mixture (Sigma-Aldrich) was added to the resulting solution and the organic and aqueous phases were separated during a centrifugation step using Phase Lock Gels (5 Prime). The DNA present in the aqueous phase was next precipitated using ethanol absolute and sodium acetate by incubating the sample at -80°C for at least an hour. The ethanol was added at a 3:1 volume ratio to the aqueous phase and sodium acetate at 0.1:1 ratio. The DNA was then spun down for 30 minutes at maximum speed in a 4°C centrifuge and the pellet was washed and spun down for 10 minutes at maximum speed twice in cold 70% ethanol. All pellets were resuspended and stored in nuclease-free water at -20°C. DNA concentration and purity were assessed using a NanoPhotometer (Implen) according to manufacturer's instructions.

2.5.2 Dual DNA and RNA extraction

DNA and RNA were extracted simultaneously form the same sample using the ZR-*Duet* DNA/RNA MiniPrep kit (Zymo Research) according to manufacturer's instructions. Both DNA and RNA were eluted from the spin column in nuclease-free water. The resulting DNA was stored at -20°C and RNA at -80°C.

2.5.3 DNA degradation

The DNA samples were digested into singe nucleotides using the DNA Degradase Plus enzyme (Zymo Research). Unless otherwise specified, 1 µg of DNA was incubated with

5 U of the enzyme for 4 hours at 37°C followed by inactivation of the Degradase enzyme at 70°C for 20 minutes. The samples were diluted 1:8 in LC/MS-grade water following the addition of equimolar internal standard mix to a final concentration of 250 nM and analysed on the mass spectrometer. The internal standard mix consisted of the following isotopically labelled nucleosides: 2'-deoxycytidine (^{15}N3) (Cambridge Isotope Laboratories), 5-methyl-2'-deoxycytidine-d3 (Toronto Research Chemicals) and 5-hydroxymethyl-2'-deoxycytidine-d3 (Toronto Research Chemicals).

2.5.4 Liquid chromatography and mass spectrometry analysis on the MaXis QTOF mass spectrometer

The QTOF-UHPLC analysis was conducted using a MaXis HD quadrupole electrospray time-of-flight (ESI-QTOF) mass spectrometer (Bruker Daltonik GmbH), which was coupled to an Ultimate 3000 UHPLC (Thermo Fisher Scientific). Analyses were performed using ESI positive-ion mode. The capillary voltage was set to 4500 V, nebulizing gas at 4 bars, drying gas at 12 L/min at 220°C. The TOF scan range was from 50 – 600 mass-to-charge ratios (m/z). Liquid chromatography was performed using a Waters Acquity UPLC HSS T3, 1.8 µM, 2.1 x 50 mm reverse phase column (Waters) with a flow rate of 0.3 mL/min, and an injection volume of 15 µL. Mobile phases A and B consisted MS-grade water (Merck), and 0.1% formic acid in acetonitrile (Merck), respectively. Gradient elution was performed as follows; 0.0 – 2.0 min, 1.0% B; 2.0 - 5.0 min, 20% B; 5.0 – 6.0 min, 95% B; 6.0 – 8.0 min, 95% B; 8.0 – 8.1 min, 1% B; 8.1 – 10.0 min, 1% B.

The MS instrument was calibrated using a range of sodium formate clusters introduced by 10 µL loop-injection prior to the chromatographic run. The mass calibrant solution consisted of 3 parts of 1 M NaOH to 97 parts of 50:50 water:isopropanol with 2% formic acid. The observed mass and isotope pattern matched the corresponding theoretical values as calculated from the expected elemental formula within 2 ppm mass accuracy. 2'-deoxycytidine, 5'-methyl-2'-deoxycytidine and 5'-hydroxymethyl-2'-deoxycytidine were detected as [M + H]+, [M + Na]+ and [2M + H]+ ions with within 0.01 Da. Data processing was performed using the Compass Data Analysis software version 4.3 (Bruker Daltonik GmbH). The work involving the MaXis HD ESI-QTOF mass spectrometer was performed with the assistance of Dr Shaun Reeksting from the

Material and Chemical Characterisation Facility at the University of Bath.

2.5.5 Liquid chromatography and mass spectrometry analysis on the Xevo triple quadrupole mass spectrometer and the validation of the in-house protocol against a previously published external method

The mass spectrometry analysis was conducted using a Xevo TqD triple quadrupole mass spectrometer (Waters), which was coupled to a Waters Acquity H-class UPLC (Waters). The capillary voltage was set to 2750 V and the pressure limits at between 4 bar and 600 bar. The source temperature was set at 150°C and the desolvation temperature at 500°C. The source gas flow was set at 1000 L/h for desolvation and at 100 L/h for the cone. The points-per-peak frequency was set at 30.5 with the run time set at 12 minutes. Argon was used as the collision gas. The instrument was tuned using an equimolar mixture of synthetic standards 2′-deoxycytidine (Sigma-Aldrich), 5-methyl-2′-deoxycytidine (Cambridge Bioscience) and 5-hydroxymethyl-2′-deoxycytidine (Cambridge Bioscience). All samples and calibration curves were spiked with an equimolar 250 nM mixture of isotopically labelled internal standards: 2'-Deoxycytidine (^{15}N3) (Cambridge Isotope Laboratories), 5-Methyl-2'-deoxycytidine-d3 (Toronto Research Chemicals) and -(Hydroxymethyl)-2'-deoxycytidine-d3 (Toronto Research Chemicals). The mass spectrometer was operated in multiple reaction monitoring (MRM) mode with the dwell time set to 0.038 seconds for all the MRM transitions.

Liquid chromatography was performed using a Waters Acquity UPLC HSS T3, 1.8 µM, 2.1 x 50 mm reverse phase column (Waters) or using a Kinetex 1.7µM, 2.1 x 100 mm HILIC column (Phenomenex) with flow rates between 0.1 and 0.3 mL/min, and an injection volume of 20 µL. Mobile phases A and B consisted MS-grade water (Merck), and 0.1% formic acid in acetonitrile (Merck), respectively. Gradient elution was performed as follows; 0.0 - 5.0 min, 100% A; 5.0 - 7.0 min, 100% B; 7.1 - 12.0 min, 100% A. Data processing was performed using the TargetLynx software (Waters). The work involving the Xevo TqD triple quadrupole mass spectrometer was performed with the assistance from Dr Jack Rice and Andrew Kannan, under the supervision of Professor Barbara Kasprzyk-Hordern from the Department of Chemistry at the University of Bath.

The limits of detection (LOD) and limits of quantification (LOQ) for both mass spectrometers are defined here as a lowest analyte concentration that gives a signal to noise ratio above 3 and 10, respectively. The results were obtained from three independent synthetic standard calibration curves prepared on the QTOF instrument and six synthetic standard calibration curves prepared on the triple quadrupole mass spectrometer. The LOD and LOQ values given for both mass spectrometers represent the lowest values experimentally obtained to date on each mass spectrometer. The linearity of calibration curves was determined by plotting the concentrations of the standards used in the dilution curves against the relative response obtained on the mass spectrometer. The Excel software (Microsoft) was used to calculate the R^2 values of the lines of best fit, which denote the linearity across the range of the concentrations tested. The ranges of concentrations used to prepare the calibration curves were pre-determined experimentally based on the range of responses observed in CRC cell lines DNA samples.

The intraday instrument validation was conducted by calculating the mean relative response obtained from three injections of the same technical triplicate of each quality control (QC) standard injected at three different times during the same day. The percentage deviation, which measures the reproducibility, between the three injections was calculated by dividing the standard deviation by the mean relative response value and converting the resulting number into percentage.

The interday instrument validation was conducted by calculating the mean relative response obtained from analysis of the freshly prepared technical triplicate of each QC standard injected on six different days. The percentage deviation, which measures the reproducibility, between the six injections was calculated by dividing the standard deviation by the mean relative response value and converting the resulting number into percentage.

The validation of the in-house protocol against a previously published external method was conducted through a collaboration with Dr Anna Caldwell and Dr Alison Brewer from King's College London (KCL). The samples were analysed on the mass spectrometer by Dr Anna Caldwell as detailed in their previously published method (Burr et al., 2018).

It should be noted that in all the figures depicting mass spectrometry results in Chapter 3, the compounds analysed are denoted as “dC” (2’-deoxycytidine), “5mdC” (5-methyl-2’-deoxycytidine) and “5hmdC” (5-hydroxymethyl-2’-deoxycytidine). This is due to the mass spectrometer analysing these compounds as nucleosides, rather than nucleotides. To avoid confusion, however, they are referred to in the main text and subsequent Chapters as “cytosine”, “5mC” and “5hmC”.

2.5.6 Dot blots

The DNA, either 4 µg/sample or 2 µg/sample, was diluted in nuclease-free water and sodium hydroxide to give a 0.1 mM final NaOH concentration in a total volume of 50 µL. These reactions were placed at 95°C for 5 minutes in order to denature the DNA. Following this incubation, an equal volume of cold 2 M ammonium acetate solution was added to each sample. For each experiment a set of short double-stranded DNA standards consisting entirely of cytosine,5- methylcytosine or 5-hydroxymethylcytosine (Zymo Research) was included. These were used to confirm the specificity of the antibodies utilized and the following amounts of each standards were used for every DNA denaturation reaction: 10 ng of the cytosine standard, 5 ng of the 5-methylcytosine standard and 0.125 ng of the 5-hydroxymethylcytosine standard. The quantities used of all three DNA sequences have been determined experimentally to give the optimal staining strength.

Each DNA reaction was then serially diluted to yield three separate technical replicates. In the case of 4 µg/sample the following amounts of DNA for each sample were spotted onto the membrane: 2 µg, 1 µg and 0.5 µg; while in the case of 2 µg/sample 1 µg, 0.5 µg and 0.25 µg were added. Following that, a rectangular piece of the Amersham Hybond-N+ nitrocellulose membrane (GE Healthcare) was incubated for 10 minutes at RT on the rocker in 6X saline sodium citrate (SSC) solution.

The pre-wet membrane was placed on top of a filter paper of the same size and inserted into the Minifold Dot Blot System (GE Healthcare). Samples were carefully spotted onto the membrane using the well system provided by the Minifold apparatus. All the empty wells were filled with an equal volume of 6X SCC solution to ensure that the subsequent vacuum suction was applied evenly throughout the membrane. Following

the solution from all the wells passing through the membrane under the applied vacuum, the membrane was allowed to air dry for 15 minutes before UV fixing it at 700 mJ/cm^2 using the CL-10000 Series UV Crosslinker (Ultra-Violet Products).

The membrane was incubated in the blocking buffer consisting of 5% (w/v) milk (Marvel) and 0.1% (v/v) Tween-20 (Sigma-Aldrich) dissolved in PBS for 1.5 hours at RT before an overnight incubation with the anti-5hmC primary antibody (Active Motif) diluted in blocking buffer. Following that, the membrane was washed three times in PBS-Tween buffer before an incubation for 1 hour at RT with an anti-rabbit secondary antibody (Abcam) diluted in blocking buffer. After three more washes of the membrane, the Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific) was used according to manufacturer's instructions to allow chemiluminescence-based detection of horseradish peroxidase signal. The signal was visualised using the Fusion SL imaging system (Vilber Lourmat).

Following the visualisation, the membrane was washed three times in PBS-Tween buffer. It was then incubated twice in the Restore PLUS stripping buffer (Thermo Fisher Scientific) for 20 minutes at 37°C to remove the antibodies before being washed three times in PBS-Tween buffer again. The membrane was then blocked and probed with the antibodies again in the same way as described above; this time with an anti-5mC primary antibody (Merck Millipore) and an anti-mouse secondary antibody (Sigma-Aldrich). The details of the antibodies used for all dot blot experiments are given in Table 2.4.

The total DNA quantity spotted in each well was assessed by methylene blue staining. The membrane was stained in 0.05% (w/v) methylene blue solution prepared in 0.3 M sodium acetate for 1 hour at RT. The staining was visualised using ChemiDoc-It^2 UV Imager (Ultra-Violet Products). All staining results were analyses using the ImageJ software.

Table 2.4 The details of the antibodies used for dot blot experiments

Antibody name	Catalogue number	Manufacturer	Antibody type	Species of origin	Dilution
Anti-5-hydroxy methylcytosine	39791	Active Motif	Primary	Rabbit	1:1,000
Anti-5-methyl cytosine	MABE146	Merck Millipore	Primary	Mouse	1:2,000
Anti-Mouse IgG (Fab specific)	A2304	Sigma-Aldrich	Secondary	Goat	1:2,000
Anti-Rabbit IgG	ab205718	Abcam	Secondary	Goat	1:2,000

2.6 Assessment of transcript levels

2.6.1 RNA extraction

The RNA was extracted from harvested cells using the GeneJet RNA Purification Kit (Thermo Fisher Scientific) according to manufacturer's instructions except for an additional on-column DNase I digestion step using the RNase-free DNase Set (Qiagen). Briefly, 350 µL of Wash Buffer 1 was added to the purification column and centrifuged for 1 min at 12,000 x g. The 10 µL of DNase I stock solution was added to 70 µL of Buffer RDD and gently mixed. This mixture was pipetted onto the membrane of the column and incubated for 15 minutes at RT. Lastly, 350 µL of Wash Buffer 1 was added to the column and centrifuged for 15 sec at 12,000 x g.

Alternatively, RNA was extracted using the TRI reagent (Sigma-Aldrich) according to manufacturer's instructions with an additional wash of the RNA pellet in 75% (v/v) ethanol. All RNA samples were eluted or dissolved in nuclease-free water and stored at -80°C. The RNA concentration and purity were assessed using a NanoPhotometer (Implen) according to manufacturer's instructions.

2.6.2 Reverse transcription

The reverse transcription reactions were performed using the QuantiTect Reverse Transcription Kit (Qiagen) according to manufacturer's instructions using 1 µg of RNA unless otherwise stated. For each set of RNA samples being reverse transcribed at the same time at least one "no reverse transcriptase" (no RT) was included where the RT enzyme was replaced with nuclease-free water. All reactions were diluted to a final 2.5 ng/µL cDNA concentration in nuclease free water and stored at -20°C.

2.6.3 Quantitative real-time polymerase chain reaction

The qRT-PCR reactions were set up in MicroAmp Fast 96-well Reaction Plate (Applied Biosystems) using the Fast SYBR Green Master Mix (Applied Biosystems) according to manufacturer's instructions. For each reaction, 5 ng of cDNA was used in a 10 µL final volume with 200 nM final primer concentration unless otherwise specified. All the reactions were performed on the StepOne Real-Time PCR machine (Applied Biosystems). All primer pairs were tested *in silico* for off target amplification using the NCBI primer design tool (National Center for Biotechnology Information). The excessive hairpin and primer dimer formation was assessed *in silico* through the OligoAnalyzer Tool (Integrated DNA Technologies). A "no RT" and "water-only" control was included for each set of samples reverse transcribed at the same time and each primer pair used, respectively. Unless otherwise specified, all of the results were analysed using the "Delta-delta Ct" method (Livak and Schmittgen, 2001).

The assessment of *TET* transcript levels in a panel of CRC cell lines was performed by Tracey Collard (University of Bristol) using primers designed by the author.

2.6.4 Selection of the most stably expressed reference genes - the GeNorm method

The stability of reference genes in the SW480 and SW620 cell lines was assessed using the geNorm 12 gene kit (Primerdesign). The kit's protocol was modified to include the qRT-PCR running conditions specified in section 2.6.3. with the exception of primer concentration, which was the same as in the manufacturer's instructions. The results of this experiment were analysed using the qBASE+ software (Biogazelle).

Table 2.5 Details of human and mouse primers used in the qRT-PCR experiments
The reference genes are marked with an asterisk (*)

Target	Forward primer sequence (5'→3')	Reverse primer sequence (5'→3')	Source
TET1	CAGATTAGTCAGGAAGGAAGATGTAA	ATTTTCCAGGGCTTAAAGTCTTGA	Primerdesign
TET2	GCAGCACACCCTCTCAAGATT	AATTCAGCAGCTCAGTCCCTTACT	NCBI primer design tool
TET3	AGAACCAGGTGACCAACGAG	CGCAGCGATTGTCTTCCTTG	NCBI primer design tool
OCT4	GAAAGCGAACCAGTATCGAGAAC	CCCCTGAGAAAGGAGACCCA	(Zhao et al., 2015)
ALB	CTTGAATGTGCTGATGACAGG	GCAAGTCAGCAGGCATCTCAT	(Sirico et al., 2012)
*RPS13**	CGAAAGCATCTTGAGAGGAACA	TCGAGCCAAACGGTGAATC	(Jacob et al., 2013)
*YWHAZ**	ATGCAACCAACACATCCTATC	GCATTATTAGCGTGCTGTCTT	(Erkens et al., 2006)
Tet1	TGAAGATGACAAGCAAACC	TTGTTGAGCGGAAGGTGTGT	Primerdesign
Tet2	GACTCAACGGTTATCAGGCTTTT	CATTTGCTCTTTATTCTTCCTCTGTAA	Primerdesign
Tet3	CTCCCCTGCTGTCTTCAGA	CCTGAGGCTCTGTGGAAGTA	Primerdesign
*Polr2a**	CCGGATGAATTGAAGCGGATGT	CCTGCCGTGGATCCATTAGTCC	(Castro et al., 2018)

Target	Forward primer sequence (5'→3')	Reverse primer sequence (5'→3')	Source
Nfe2l2	CCAGCACATCCAGACAGACA	TATCCAGGGCAAGCGACTCA	NCBI primer design tool
Klf4	AACCGTTGGCGTGAGGAAC	AGTCTAGGTCCAGGAGGTCG	NCBI primer design tool
Dusp6	GGAAGTGTGTGTATGTGTGCGAGTA	CCTTCCTCTCAACCTGTTGCA	(Lin et al., 2010)
Eps8	AAGAACAGAGGCGTGAGAGC	TGGTGAAGATCCGTAACCATTCA	NCBI primer design tool
Ihh	CCTCAGACCGTGACCGAAAT	GCCGAATGCTCAGACTTGAC	NCBI primer design tool
Idh1	ATGCAAGGAGATGAAATGACACG	GCATCACGATTCTCTATGCCTAA	(Itsumi et al., 2015)
Ogdh	CCCTTTCCCTGAGTCGAAGC	ACATGGTGCCCTCGTATCTGA	NCBI primer design tool
Mdh1	ATGATGGGTGTTCTGGACG	TCACATTGGCTTTCAGTAGG	(Suhara et al., 2015)
Mdh2	CCAGTGAACTCCACCATCCC	AACCCTTTAGCTCTGCCACAA	NCBI primer design tool
Aco1	AACACCAGCAATCCATCCGT	GGTGACCACTCCACTTCCAG	(Hochberg et al., 2015)
Aco2	TCTCTAACAACCTGCTCATCGG	TCATCTCCAATCACCACCCACC	(Villena et al., 2007)
Gpi1	TTGTCCCCTGAGACTTCCCT	ACTTTGGCCGTGTTCGTAGA	NCBI primer design tool

Table 2.6 Details of all human reference gene primers from geNorm 12 gene kit used in the qRT-PCR experiments
The nucleotide sequences of these primers are a property of Primerdesign and are hence not provided

Target	Accession number	Anchor nucleotide
18S	M10098	235
ACTB	NM_001101	1194
GAPDH	NM_002046	1089
RPL13A	NM_001135	727
UBC	NM_021009	271
B2M	NM_004048	302
YWHAZ	NM_003406	2534
CYC1	NM_001916	890
EIF4A2	NM_001967	876
SDHA	NM_004168	1032
TOP1	NM_003286	2362
ATP5B	NM_001686	1200

2.6.5. Primer efficiency analysis

To assess the qRT-PCR primer efficiency, selected primers were ran with decreasing amounts of the same cDNA. The cDNA was serially diluted 1:2 with the starting amount per well set at 10 ng and final amount per well equal to 0.625 ng. All the reactions were performed exactly as detailed in section 2.6.3. The Pfaffl equation was used to correct the qRT-PCR results for primer efficiency (Pfaffl, 2001).

2.7 Absolute transcript quantification

2.7.1 cDNA amplification

The same primers targeting the three *TET* transcripts in qRT-PCR reactions were used in a regular PCR reaction using cDNA previously prepared from HCT116 RNA. The OneTaq DNA polymerase (New England Biolabs) was utilized for this amplification and the reactions were performed according to manufacturer's instructions. The same primer concentration and the amount of cDNA per reaction as in all qRT-PCR experiments was used (see section 2.6.3).

2.7.2 Agarose gel electrophoresis

The 1.2% - 2% agarose (Sigma-Aldrich) gels were stained using the GelRed Nucleic Acid Gel Stain (Cambridge Bioscience). The samples were loaded using the BlueJuice loading buffer (Thermo Fisher Scientific). The 100 bp DNA ladder (Invitrogen) or 1 kbp DNA ladder (New England Biolabs) were used to assess the size of the DNA bands. The gels were visualized using the ChemiDoc-It^2 UV Imager (Ultra-Violet Products).

2.7.3 PCR product clean up from agarose gels

The agarose gels were illuminated with UV light with care being taken to minimize the duration of exposure. The bands containing the PCR products were cut out of the gel using a scalpel blade. The DNA was then extracted from the agarose gel fragments using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) according to manufacturer's instructions.

2.7.4 Ligation of PCR products into a plasmid vector

The purified PCR products were ligated into a pGEM-T Easy Vector System (Promega Corporation) according to manufacturer's instructions. An overnight ligation at 4°C was utilized to maximise the number of transformants.

2.7.5 Bacterial transformation

The plasmid vectors were transformed into in-house made supercompetent E. coli using the following protocol. The E.coli were thawed on ice and 1 - 2 µL of the ligation reaction was added to the bacteria. The mixture was incubated on ice for 12 minutes before heat shocking them at 42°C for 45 seconds. Following that the mixture was incubated for 2 minutes on ice and pre-warmed Super Optimal broth with Catabolite repression (SOC) media was added at a 3:1 volume ratio of SOC media to bacteria and plasmid mixture. The tubes were placed in a shaking incubator at 225 rpm and 37°C for 1 hour. The liquid was then streaked on Luria-Bertani (LB) agar plates supplemented with 100 µg/mL carbenicillin, 50 µg/mL 5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside (X-Gal) and 100 µM isopropyl-β-D-thiogalactoside (IPTG) and incubated at 37°C overnight.

2.7.6 Bacterial miniprep

The white colonies from plates for each plasmid construct were picked using a pipette tip for plasmid extraction. The miniprep was performed using the Zyppy Plasmid Miniprep Kit (Zymo Research) according to manufacturer's instructions. The plasmid DNA was eluted in nuclease-free water and stored at -20°C.

2.7.7 Restriction digest confirmation of insert ligation

The purified plasmid DNA was digested with the EcoRI enzyme (New England BioLabs) according to manufacturer's instructions. To confirm the presence of the insert, the digestion reactions were run on an agarose gel as described in section 2.7.2.

2.7.8 Preparation of standard curves using the ligated plasmid constructs

The concentration of the purified DNA constructs was assessed by taking ten independent readings using a NanoPhotometer (Implen) according to manufacturer's instructions. The average of the ten readings for each of the three plasmids was taken as the plasmid concentration. A series of six five-fold dilutions of pre-assessed

concentrations of each of the three constructs was ran alongside the regular qRT-PCR reactions, as described in section 2.6.3. The absolute levels of each of the *TET* transcripts were calculated by using the equations of the three standard dilution curves to convert the average Ct values obtained for each *TET* transcript in the three cell line samples to picograms of vector used in the reactions.

2.8 Protein levels and activity analysis

2.8.1 Protein extraction

The harvested and pelleted cells were re-suspended in radioimmunoprecipitation assay (RIPA) buffer supplemented with cOmplete, EDTA-free Protease Inhibitor (PI) Cocktail (Roche) to a final 1X concentration. The RIPA buffer was prepared fresh before each experiment and consisted of 150 nM sodium chloride, 1% (v/v) IGEPAL CA-630 (Sigma-Aldrich), 0.5% (w/v) sodium deoxycholate, 0.1% (w/v) sodium dodecyl sulphate (SDS) and 50 mM Tris pH 8.0. The extraction reactions were then rotated for 10 minutes at 4°C on a roller before being stored at -80°C prior to quantification.

2.8.2 Nuclear and cytoplasmic protein extraction

The nuclear and cytoplasmic fractions of cells were isolated using the NE-PER Nuclear and Cytoplasmic Extraction Reagents kit (Thermo Fisher Scientific) according to manufacturer's instructions. The cOmplete, EDTA-free Protease Inhibitor Cocktail (Roche) was added to CER I and NER reagents to a final 1X concentration immediately before use. The cytoplasmic and nuclear extracts were subsequently stored at -80°C.

2.8.3 Protein quantification

The protein concentration in the extracts was assessed using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) according to manufacturer's instructions. The absorbance (A562 nm) was measured using the plate reader (Turner Biosystems). The

protein quantification experiments were designed by the author and the experimental procedures were performed and analysed by Dr Jennifer Pinnell.

2.8.4 Western blotting

The protein samples were diluted in the RIPA + PI buffer to a final volume between 15 and 25 µL containing 30 µg of protein in total. An equal amount of Laemmli sample buffer (Sigma-Aldrich) was added to the diluted samples and the resulting mixture was incubated at 95°C for 5 minutes. The samples were loaded into the wells of Novex WedgeWell 4-12% Tris-Glycine Mini Gels (Thermo Fisher Scientific) placed inside the XCell *SureLock* Mini-Cell (Invitrogen) according to manufacturer's instructions. The Spectra Multicolour Broad Range Protein Ladder (Thermo Fisher Scientific) was used to provide a size reference for the protein bands. The gels were ran for 90 minutes at 125 V in running buffer consisting of 0.3% (w/v) Tris base, 1.5% (w/v) glycine and 0.1% (w/v) SDS solution.

The transfer of proteins from the gel to the PVDF membrane (GE Healthcare) was carried out inside the XCell II Blot Vertical Module (Invitrogen) according to manufacturer's instructions. The transfer was carried out for 90 minutes at 25 V in transfer buffer consisting of 0.3% (w/v) Tris base, 1.5% (w/v) glycine and 20% (v/v) methanol. The membrane was blocked overnight at 4°C in blocking buffer consisting of 5% (w/v) milk (Marvel) and 0.1% (v/v) Tween-20 (Sigma-Aldrich) dissolved in PBS. The following day the membrane was incubated with the primary antibodies diluted in blocking buffer for 2 hours at RT. The membrane was then washed three times in PBS-Tween buffer before a 1 hour incubation at RT with the secondary antibodies diluted in blocking buffer. The details of all antibodies used for Western blotting experiments are given in Table 2.7.

After three more washes of the membrane in PBS-Tween, the Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific) was used according to manufacturer's instructions to allow chemiluminescence-based detection of horseradish peroxidase signal. The signal was visualised using the Fusion SL imaging system (Vilber Lourmat). Following the visualisation, the membrane was washed three times in PBS-Tween buffer. It was then incubated twice in the Restore PLUS stripping buffer (Thermo Fisher

Scientific) for 20 minutes at 37°C to remove the antibodies before being washed three times in PBS-Tween buffer again. The membrane was then blocked and probed with antibodies again in the same way as described above. The intensity of the bands was measured using the ImageJ software and the results were normalized to the β-tubulin signal. The Western blotting assays were designed by the author and the experiments were performed and analysed by Dr Jennifer Pinnell.

Table 2.7 The details of the antibodies used for Western Blot experiments

Antibody name	Catalogue number	Manufacturer	Antibody type	Species of origin	Dilution
Monoclonal anti-TET1 (GT1462)	MA5-16312	Invitrogen	Primary	Mouse	1:1,000
Monoclonal anti-TET2 (hT2H 21F11)	MABE462	Sigma-Aldrich	Primary	Mouse	1:1,000
Polyclonal anti-TET3 [C3], C-term	GTX121453	GeneTex	Primary	Rabbit	1:1,000
Monoclonal anti-β-Tubulin	T0198	Sigma-Aldrich	Primary	Mouse	1:1,000
Anti-Mouse IgG (Fab-specific)	A2304	Sigma-Aldrich	Secondary	Goat	1:10,000
Anti-Rabbit IgG	ab205718	Abcam	Secondary	Goat	1:10,000

2.8.5 TET enzymes activity assay

The activity of the TET enzymes in cell line samples was assessed using the TET Hydroxylase Activity Quantification Kit (Abcam) according to manufacturer's instructions. The absorbance (A450 nm and A655 nM) was measured using the plate reader (Turner Biosystems). The TET activity assay was designed by the author and the experiments were performed and analysed by Dr Jennifer Pinnell.

2.9 Statistical analysis and figure preparation

The statistical significance of the difference observed between two experimental groups was determined by two-tailed, unpaired student's t-test unless otherwise specified.

The statistical significance of the difference observed between three or more experimental groups was determined by a one-way ANOVA followed by the Tukey's single-step multiple comparison unless otherwise specified.

Unless otherwise specified, all figures were prepared using the Excel software (Microsoft) and all tables were made using the Word software (Microsoft).

Figure 1.1, Figure 2.1, Figure 3.1, Figure 3.2, were prepared using the PowerPoint software (Microsoft).

Figure 1.2 was modified from Servier Medical Art, licensed under a Creative Common Attribution 3.0 Unported License using the PowerPoint software (Microsoft).

Figure 1.3 and Figure 3.3 were made using the ChemDraw software (PerkinElmer).

Chapter 3

Establishing an in-house mass spectrometry detection method for global 5-methylcytosine and 5-hydroxymethylcytosine levels

3.1 Chapter summary

The aim of the work described in this chapter was to develop and optimise an in-house mass spectrometry method for detection of global 5mC and 5hmC levels. The rationale behind this project was that having the ability to analyse DNA samples at University of Bath would be both time- and cost-effective.

The MaXis HD ESI-QTOF and Xevo TqD triple quadrupole mass spectrometers were assessed for their sensitivity of 5mC and 5hmC detection. This was followed by the optimisation of the sample preparation process, improvement of the liquid chromatography performance, assessment of the inter- and intraday variability of the method validation against a previously published mass spectrometry protocol at an external institution. In summary, a robust in-house mass spectrometry-based method of cytosine modifications was developed with limits of detection as low as 30 attomoles of the analyte.

Lastly, the results are presented in the context of the wider literature and are discussed in relation to the objectives set out prior to the start of the experiment series with the limitations of the in-house method and possible future improvements also described.

3.2 Introduction

The rise of the field of epigenetics was made possible by the various methods of detection and quantification of DNA and histone modifications. For the cytosine modifications 5mC and 5hmC, these can be categorised as assessing either their global or locus-specific levels (comprehensively reviewed by Qing et al., 2017). Briefly, the latter category includes methods such as bisulfite and oxidative bisulfite conversion, DNA immunoprecipitation with 5mC- and 5hmC-specific antibodies and T4 bacteriophage β-glucosyltransferase based glucosylation of 5hmC. The methods used to investigate the global levels of 5mC and its oxidative derivatives include, among others, high performance liquid chromatography, thin layer chromatography, dot blots as well as immunohistochemistry and immunofluorescence. The locus-specific changes in 5mC and 5hmC levels affect the gene and enhancer activity. The techniques used to measure them are hence often coupled with gene expression measurements such as RNA-seq. This allows one to better understand how these epigenetic alterations affect the expression of a particular gene or multiple loci at the same time. The measurement of global 5mC and 5hmC levels enables one to gain insight into whole-genome changes of these cytosine modifications. This is particularly useful for understanding the dynamics of cytosine modifications turnover and their role during biological processes such as vertebrate development, cellular differentiation and cancer progression. A prominent example of this is the finding that global 5hmC levels are reduced in numerous cancer types as compared to healthy tissues of the same origin and have hence been proposed as a potential diagnostic tool (Chen et al., 2017; Gilat et al., 2017).

Since the discovery of the TET enzymes and confirmation of the presence of 5hmC in human tissues, liquid chromatography coupled to mass spectrometry (LC-MS) has emerged as a gold standard for quantifying the global levels of 5mC and 5hmC levels initially, followed by global levels of 5fC and 5caC shortly after (Münzel et al., 2010; Ito et al., 2011; Bachman et al., 2014; Bachman et al., 2015). This method was initially pioneered by Thomas Carell's group and subsequently became widely used in the epigenetic field with a wide variety of sample preparation methods and LC-MS running conditions published to date (Bachman et al., 2014; Amouroux et al., 2016; Thienpont et al., 2016; Burr et al., 2018). This is due to its established superior sensitivity and the recent advancements allowing the use of very low quantities of DNA for the analysis

(Le et al., 2011; Yuan et al., 2020). The method is also more time efficient and allows processing of considerably more samples in one run as compared to antibody-based detection of 5hmC levels such as dot blots. From studying these protocols, it becomes clear that there are a number of steps that have to be optimised in order to create a procedure that is best suited for the in-house mass spectrometry instruments available at the University of Bath. This is particularly important because of the requirement for high sensitivity to reliably detect 5hmC in biological samples due to its low abundance in genomic DNA. These steps include the selection of the appropriate mass spectrometer for this type of analysis, optimisation of the DNA extraction and digestion steps, selection of the correct liquid chromatography column and running conditions as well as fine-tuning the running parameters on the mass spectrometer.

3.2.1 Liquid chromatography

Liquid chromatography coupled to mass spectrometry is an analytical technique characterised by very high sensitivity and specificity that is widely used in, amongst others, proteomic and metabolomic analysis experiments. It allows one to identify unknown compounds in a mixture and predict their structure based on the mass and charge of the fragment ions. While historically LC instruments relied on the force of gravity to achieve separation of compounds in a mixture, the development of high-performance liquid chromatography (HPLC) led to considerably improved resolution as well as reduced sample processing time. HPLC relies on pumps to generate a highly pressurised liquid stream that passes through the HPLC column where the separation of a compound mixture takes place. This occurs because different molecules exhibit varying affinities for the solid phase that makes up the inside of the column due to various degrees of their polarity (Figure 3.1).

In this particular method, reverse phase chromatographic separation of analytes has been utilised. It is the most commonly used chromatography type in HPLC analysis, and it relies on the interaction of the hydrophobic nucleosides with the hydrophobic moieties that make up the stationary phase of the HPLC column (Claessens, 2001). In that scenario, the more hydrophilic compounds will elute from the column at an earlier time point following the initiation of the organic mobile phase gradient. The most

commonly used stationary phases in reverse phase HPLC are octadecyl carbon chain (C18) and octyl carbon chain (C8). These long carbon molecules are attached to, most commonly, silica particles that form the stationary phase. The C18 chains are more hydrophobic that the C8 chains and hence will lead to a longer retention of less polar analytes. Both of these molecules mainly interact with the analytes via the hydrophobic van der Waal forces (Hanai, 2017).

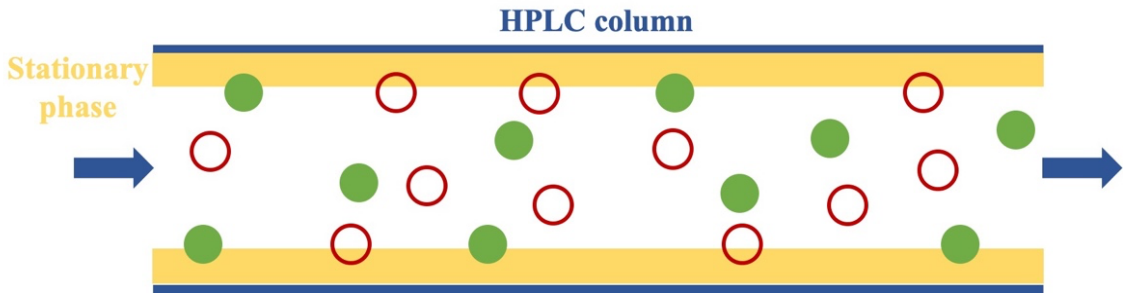


Figure 3.1 A schematic representation of passage of compounds through a HPLC column
The compounds passing through the HPLC column bind to the stationary phase and are eluted depending on their polarity and the mobile phase composition. In the case of the reverse phase HPLC, hydrophobic compounds (green, full circle) will be retained on the stationary phase longer. In normal phase HPLC, the hydrophilic molecules (red, empty circle) will stay bound to the stationary phase longer.

Normal phase HPLC operates on opposite principle to reverse phase; namely it relies on stationary phase being polar which leads to longer retention of hydrophilic compounds on the column. In this case, the mobile phase flowing through the column in non-polar - examples include hexane, chloroform and 2-propanol. This is in stark contrast to reverse phase where the mobile phase is polar and most commonly includes a mix of water and either methanol or acetonitrile. Regardless of the type of chromatography used, the mobile phase can be applied to the column using either isocratic or gradient elution. Isocratic HPLC refers to a constant composition of the mobile phase throughout the run. On the other hand, gradient elution involves gradually changing the proportions of the mobile phase constituents over the run period. This allows for increased sensitivity of the downstream detection as well as improved separation of peaks corresponding to compounds with similar retention times due to structural similarities, such as cytosine

and its modifications. All the mass spectrometry experiments performed in this chapter have hence utilised gradient elution.

An interesting case of normal phase HPLC is the hydrophilic interaction liquid chromatography (HILIC) method. It involves an improvement of retention on the column, as well as the overall response, of polar analytes while running a mobile phase that is normally used in reverse phase chromatography. It provides a viable alternative to normal and reverse phase HPLC when a mixture of polar and non-polar molecules is analysed. In particular, it is likely to improve the retention time and chromatographic peak shape of polar compounds as compared to reverse phase HPLC.

3.2.2 Mass spectrometry

Following the chromatographic separation on the HPLC column, the analytes are required to arrive at the mass spectrometer detector in a gaseous ionised state. Multiple ionisation methods have been described to date, including electrospray ionisation (ESI), atmospheric pressure chemical ionisation (APCI) and matrix-assisted laser desorption/ionization (MALDI) (Siuzdak, 2004). APCI involves desolvation of the sample droplets via a heating mechanism followed by passing next to the corona discharge needle and analyte-solvent collisions resulting in the ion formation. Corona discharge refers to an ionisation process where a fluid such as air or vaporised mixture of analytes is ionised in an electrical field created by a strongly charged electrode. MALDI on the other hand utilizes a combination of a pulsing laser beam and a protective matrix compound to ionise the analyte molecules without their destruction by the laser (Clark et al., 2013). In the experiments described in this thesis, ESI was exclusively used as an ionisation method and formic acid was added to the components of the mobile phase to aid with the process of positive ion formation.

The initial stage of ESI involves the passage of the liquid containing the analytes through a fine capillary whose tip is charged with a strong electric current (Siuzdak, 2004). This allows the formation of a droplet stream through the nebulisation process which is then subjected to both dry gas and heat in the desolvation unit to facilitate the desolvation of the droplets (Figure 3.2). The ever-decreasing size of these droplets leads to an increasing charge on their surface which is subjected to forces of repulsion

between ions of the same charge, known as the Coulomb forces. Once those forces reach a point where the surface tension of the droplets is lower than the repulsion between ions (called the Rayleigh limit), the droplets are fragmented into smaller ones, also releasing some of the ions (Banerjee and Mazumdar, 2012) It has also been suggested that the ions are expelled from the droplet via the formation of what is known as a Taylor cone (Wilm, 2011). Taylor cone is also thought to form at the end of the charged capillary and lead to the nebulisation process described above. Ionisation can be performed both in a positive or negative mode. This refers to formation of either cations or anions of the analyte molecules, respectively. All the mass spectrometry experiments described in this thesis were performed in the positive ionisation mode.

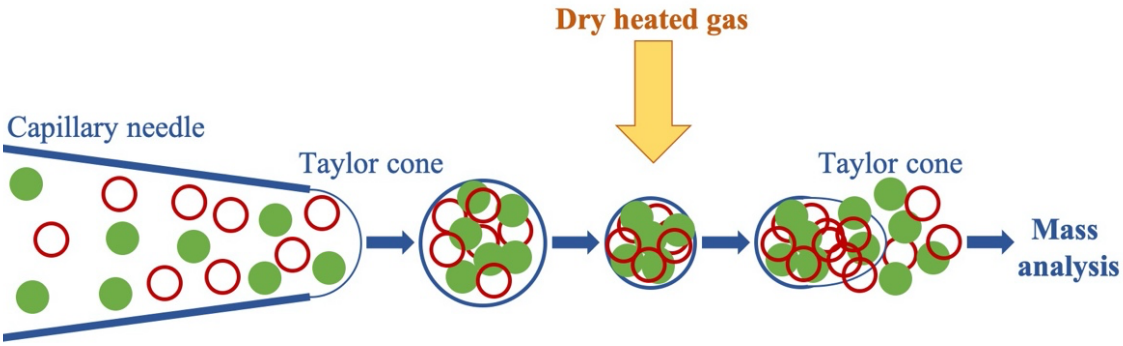


Figure 3.2 A graphical representation of the ESI process
The green, full circles represent a hydrophobic compound and the red, empty circles represent hydrophilic molecules.

Once the ions are formed by the ESI process, they are transferred into the vacuum of the mass spectrometer through the electrically charged sampling apertures. How the ions are subsequently processed and detected depends on the type of mass spectrometer being used. The two instruments used in this project are the MaXis HD quadrupole ESI-QTOF and Xevo TqD triple quadrupole mass spectrometers. They represent two different approaches to mass spectrometry: time-of-flight and triple quadrupole analysis.

Triple quadrupole mass spectrometry utilizes, as the name suggests, three quadrupoles to process the ion gas formed by the ionisation process. The quadrupoles consist of four metal cylinders arranged together and subjected to an electric field. This allows for the selection of the ions with particular mass to charge ratio at each quadrupole by varying the voltage applied to them. In a triple quadrupole system, the first and third

quadrupoles act as such “mass to charge ratio gates” for ions while the middle quadrupole forms a collision cell. The collision cell is filled with a neutral gas (argon in the case of the Xevo TqD model) which the analyte ions collide with resulting in their fragmentation.

The triple quadrupole system can be used for a number of analysis types (Pitt, 2009). These rely on utilising either the first or the third quadrupole, or indeed both (summarised in Table 3.1). In a product scan mode, the first quadrupole selects for an ion of a certain mass that is then fragmented and all the resulting daughter ions are scanned based on their mass to charge ratio in the third quadrupole and are hence all analysed on the mass spectrometer. This allows for confirmation of a structure of an unknown compound based on its structural fragmentation spectrum. In an opposite scenario, called the precursor scan mode, all ions are scanned using the first quadrupole and allowed to fragment. The third quadrupole then selects only fragments of a particular mass. This allows for identification of all the precursor ions that produce a common fragment ion. In a neural ion loss scan, both the first and the third quadrupole are used to scan for ions with a difference in mass selected for corresponding to an uncharged ion being lost in a collision cell. This allows for identification of the parent ions that loose the same neutral ion in the fragmentation process (Hoffman et al., 2006).

Table 3.1 The summary of the triple quadruple mass spectrometer running modes

Analysis type	First quadrupole	Third quadrupole	Application
Product scan	Ion selection	Scan	Structural confirmation studies
Precursor scan	Scan	Ion selection	Identification of ions producing a common fragment
Neutral loss scan	Scan	Scan (minus neutral ion mass)	Identification of ions producing a common neutral fragment
Multiple reaction monitoring	Ion selection	Ion selection	High sensitivity analysis of specific ion levels if the parent and fragment ion mass is known

The mode that is directly relevant to this project and used in all the experiments subsequently described is referred to as multiple reaction monitoring (MRM). MRM utilizes both the first and third quadrupoles to select for ions of a known mass to charge ratio. This allows one to detect only the compounds of interest thus greatly increasing the sensitivity. The main requirement of this method is that the exact mass of each precursor and daughter ion has to be known in advance or experimentally determined first.

Time-of-flight mass spectrometry refers to a method of spectrometric analysis that utilizes the measurement of the time it takes an analyte ion to reach the detector. The path of the ions is of a known distance and they are accelerated by an electric field of a known strength. The time required for the signal to be detected hence depends on the mass to charge ratio of each ion. An improvement on the regular time of flight mass spectrometers is a QTOF instrument where a quadrupole and a collision cell are introduced. The MaXis HD quadrupole ESI-QTOF is an example of such instrument. In a similar fashion to the triple quadrupole mass spectrometer, the first quadrupole is followed by a collision cell (equivalent to second quadrupole in a triple quadrupole instrument). The difference between the instruments is that the last quadrupole in the triple quadrupole mass spectrometer is replaced by the time of flight sensor in the QTOF. As such, the QTOF is more suited to studying a larger number of compounds in a mixture simultaneously. An excellent example of that is a metabolomic analysis. It is however less suited to a more targeted analysis of a few selected low-abundance compounds where sensitivity is paramount.

For the analysis of mass spectrometer data to be quantitative, a set of standards of known concentrations must be included in every sample. These are referred to as internal standards and they should be structurally closely related to the analytes of interest, so that they ionise and fragment in the same way. In the in-house method described in this chapter, a set of three such compounds was hence included in the analysis. These have an identical structure to the synthetic standards used, except for isotopically labelled moieties (see Figure 3.3 for details). A known concentration mixture of those was included in every calibration curve point and sample subsequently prepared to allow for quantitative analysis. In this way, they act in a manner similar to the reference genes in qRT-PCR experiments allowing for a quantitative sample comparison.

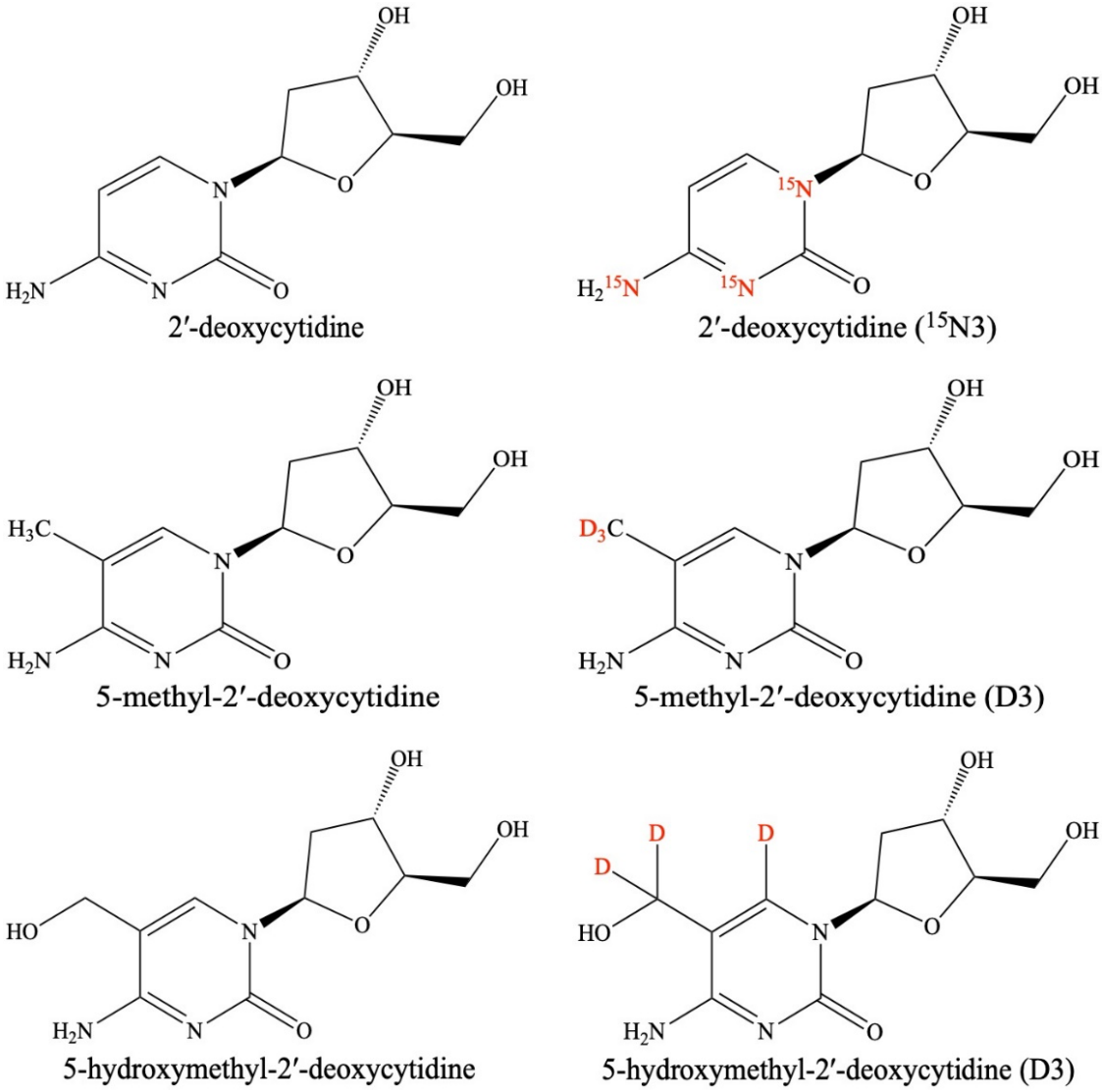


Figure 3.3 The structures of all compounds analysed on the mass spectrometer
The synthetic standards are shown on the left-hand side while the corresponding internal isotopically labelled standards are displayed on the right-hand side. The isotopically labelled moieties are shown in red.

3.2.3 The rationale behind the development of an in-house mass spectrometry method

The reduction in global levels of 5mC and 5hmC is an established characteristic of a multitude of human cancers. Importantly, the lower abundance of 5hmC in cancers' genome was proposed as a sensitive diagnostic biomarker for, amongst others, gastric, liver and colorectal tumours (Chen et al., 2013; Li et al., 2017). Mass spectrometry has

emerged as a gold standard for quantification of global levels of cytosine modifications in the DNA. It provides a superior sensitivity as compared to the methods described in the introduction to this chapter, many of which are only semi-quantitative. This analytical technique is hence crucial for this PhD project as it focuses on investigating the changes in global 5hmC levels in colorectal cancer and its metastasis as well as cellular differentiation. There are multiple advantages to setting up an in-house mass spectrometry method rather than sending the samples for analysis to an institution already performing such experiments. Firstly, by performing all the experiments in-house, one has the complete control over the whole experimental process. The necessary changes and adjustments to the protocol can be made with relative ease and the possibility of an error being made is reduced. Additionally, it is considerably more time- and cost-efficient as there are no delays and costs associated with transportation of samples. Having the infrastructure to analyse the global levels of 5mC and 5hmC in the DNA in-house allows one to run samples on the same day as they are prepared and reduce the risk associated with transport. Given the attomole sensitivity levels of the mass spectrometer, one risks skewing the results considerably if the samples or standards are subjected to degradation because of variation in storage temperatures. Lastly, it allows one to gain a deeper understanding of the mass spectrometry technique itself. By optimising all the components of the workflow to provide the required sensitivity, one gains the insight into every part of the process and experience in experiment troubleshooting. As such, the method optimisation approach outlined in this chapter could be used as a guide to establishing a network of mass spectrometry instruments at academic institutions that would allow hospitals access to quick and reliable cancer diagnosis in cases where routinely used methods, such as histological analysis, give inconclusive results. This expertise can be used to expand the method to include other compounds of interest. These might include the further oxidative derivatives of 5mC: 5-formylcytosine and 5-carboxylcytosine as well as the DNA damage marker 8-oxoguanine. The latter would be of particular interest given the fact that 5hmC colocalizes to the sites of DNA damage and 8-oxoguanine may play a role in regulating gene expression (Kafer et al., 2016; Ba and Boldogh, 2018).

3.3 Experimental objectives

Aim: to develop a cost-effective sensitive method of detection for global levels 5mC and 5hmC levels using the mass spectrometry facilities available at the University of Bath.

Objectives:

1. Determine whether the mass spectrometry facilities at the University of Bath possess the required sensitivity for robust detection of the low abundance 5hmC in tissues and cell lines
2. Optimise parameters of the sample preparation and liquid chromatography process in order to maximise the sensitivity of this method
3. Validate the in-house protocol against a previously published external method

3.4 Results

3.4.1 The assessment of the mass spectrometry facilities at the University of Bath for sensitivity and robust detection of the low abundance 5hmC in tissues and cell lines

The initial parameters for testing of the method were set up following the protocol previously used by Adele Murrell's group at the University of Cambridge (Bachman et al., 2014). These included the C18 liquid chromatography column, DNA digestion protocol and the mass spectrometer running parameters. The first mass spectrometer tested for its suitability for this type of analysis was the MaXis HD ESI-QTOF mass spectrometer. While in general TOF instruments are more widely used for an analysis of a panel of compounds simultaneously, the inclusion of the quadrupole in the MaXis instrument offers increased sensitivity that could potentially be sufficient for reliable and reproducible 5hmC levels detection.

To assess the sensitivity of this mass spectrometer, calibration curves made up of serial synthetic standard dilutions were prepared (Figure 3.4 A). The calibration curves displayed a high degree of linearity, although the R^2 values should ideally be higher than 0.999 (the generally accepted value denoting a high degree of linearity). Additionally, the high R^2 values were obtained across a narrow concentration range for both 5mC and 5hmC. The DNA extracted from whole mouse brain (known to have high 5hmC levels) and HCT116 colorectal cancer cell line (known to have low 5hmC levels) was also analysed (Figure 3.4 B). While both 5mC and 5hmC were detectable in both the mouse brain and HCT116 DNA, a large amount of digested DNA was required to do so (approximately 8 μg for the mouse brain and 30 μg for the HCT116 cell line).

Additionally, the signal obtained for 5hmC in the HCT116 DNA was close to the detection limit. The same pattern was observed across a number of experiments with different cell line samples (data not shown). In the light of these results, it became clear that the MaXis HD ESI-QTOF mass spectrometer will be unlikely to provide sufficient sensitivity for reproducible detection of global 5hmC levels in the DNA samples isolated from cancer cell lines.

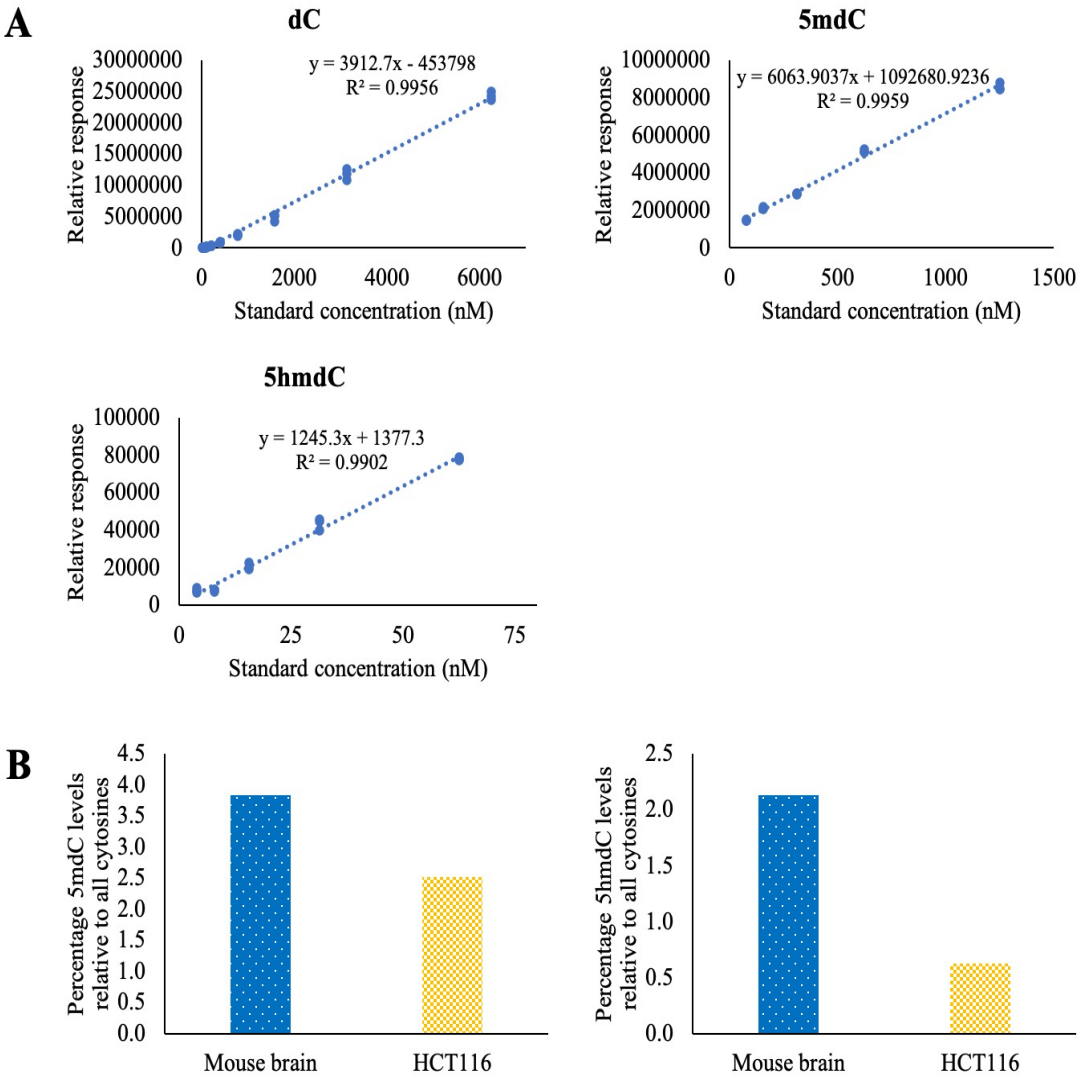


Figure 3.4 The MaXis HD ESI-QTOF mass spectrometer is capable of detecting 5hmC in DNA samples but lacks the required sensitivity

(A) Calibration curves display a high degree of linearity for all three compounds across the following ranges: dC 12 - 6250 nM: 5mdC 78 - 1250 nM; 5hmdC 4 - 63 nM. The serial dilutions were prepared using a synthetic standard mix of dC, 5mdC and 5hmdC. The R^2 values denote the degree of linearity across the concentrations analysed. The dotted line represents the line of the best fit.

(B) 5mdC and 5hmdC are detectable in DNA from mouse whole brain and HCT116 cell line. The results are expressed as percentage of total (dC+5mdC+5hmdC) cytosines measured. The results represent one replicate of an initial experiment.

To allow a direct comparison between the two mass spectrometers (Table 3.2), a similar set of calibration curves was prepared and analysed on the Xevo TqD instrument (Figure 3.5). The calibration curves showed a higher degree of linearity across a larger range of dilutions as compared to the ones prepared on the QTOF machine - the R^2 value for all three compounds was above 0.999. When directly compared, the limits of detection for all three compounds were considerably lower, almost 1000-fold for all compounds, than those obtained on the QTOF mass spectrometer (Table 3.2). The Xevo TqD triple quadrupole instrument was hence selected for subsequent method optimisation and sample analysis.

Table 3.2 The direct limits of detection comparison between the two mass spectrometers

Compound	QTOF mass spectrometer	Triple quadrupole mass spectrometer
dC	12.2 nM	22.5 pM
5mdC	0.61 nM	<1.4 pM
5hmdC	3.91 nM	<1.4 pM

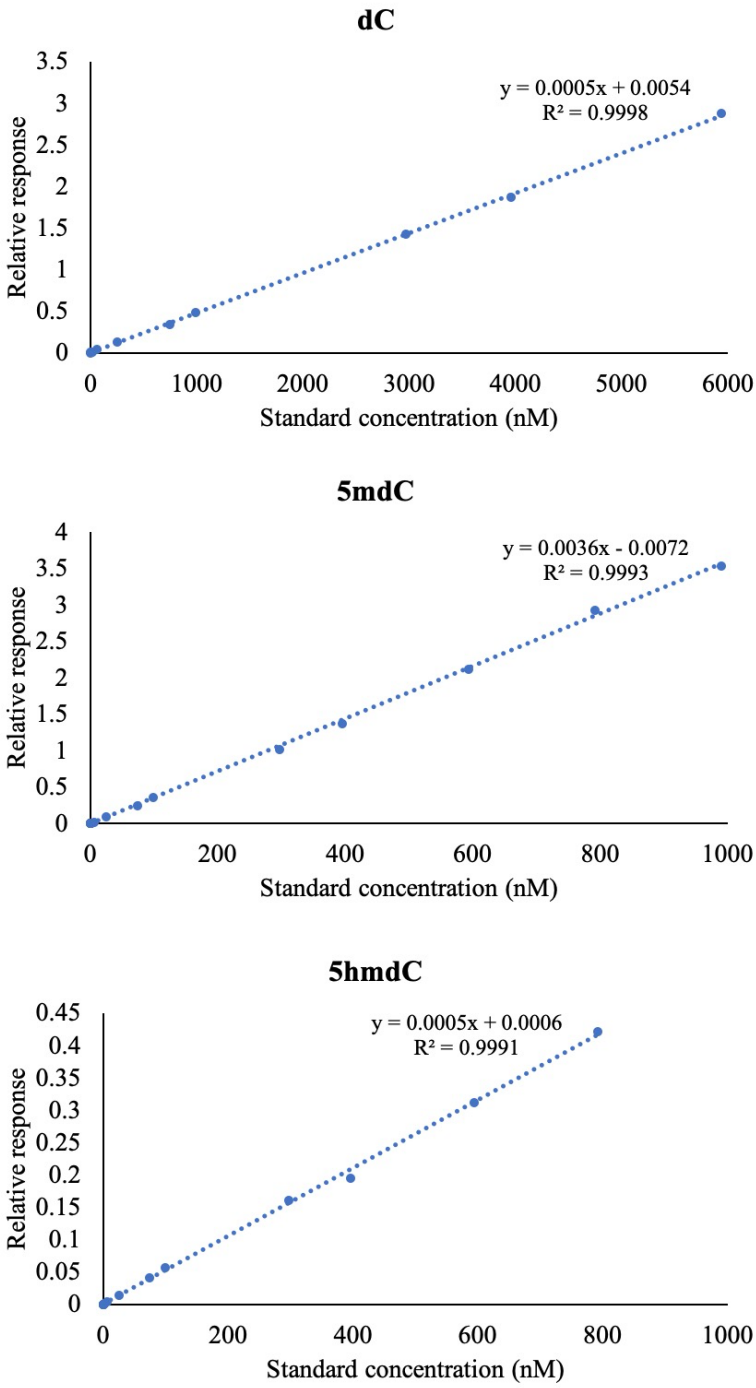


Figure 3.5 The Xevo TqD triple quadrupole mass spectrometer has a high degree of linearity across a broad 5mdC and 5hmdC concentration range

A representative example of a set of calibration curves prepared on the Xevo TqD triple quadrupole mass spectrometer. The linearity for the three compounds was obtained across the following ranges: dC 0.24 - 5940 nM (top graph); 5mdC 0.0015 - 990 nM (middle graph); 5hmdC 0.0015 - 792 nM (bottom graph). The calibration curves were prepared using a synthetic standard mix of dC, 5mdC and 5hmdC. The R^2 values denote the degree of linearity across the concentrations analysed. The dotted line represents the line of the best fit.

3.4.2 The optimisation of mass spectrometry, sample preparation and liquid chromatography parameters in order to maximise the sensitivity of the in-house protocol

The selection of the correct mass spectrometer for the in-house method development was followed by optimisation of the mass spectrometry parameters, the sample preparation and liquid chromatography set-up. This was done to maximise the analyte signal obtained and to minimise the possible effects of ion suppression. The term ion suppression refers to a reduction in the efficiency of the ionisation process. This can be caused by the presence of, for example, salts, intracellular metabolites or proteins that interfere with the nebulisation and droplet evaporation (Annesley, 2003). Presence of such contaminants can also negatively affect the lifespan of the HPLC column. It is therefore paramount to assess whether certain steps in the method workflow contribute to the suppression of the ionisation process. Additionally, given the low abundance of 5hmC in genomic DNA, the sensitivity of the in-house method described here is critical as alluded to earlier.

3.4.2.1 The optimisation of mass spectrometry parameters

The mass spectrometry conditions were optimised through tuning the mass spectrometer using a high concentration solution of each individual standard (Table 3.3). The MRM transition values for each ion, namely the mass to charge ratio of the precursor and product ions obtained from the literature (Bachman et al., 2014) were experimentally confirmed. The retention time is provided as an average of the retention times of all calibration curve points shown in Figure 3.5. The structural similarity and little difference in retention time between 5mC and 5hmC led to an addition of a 5hmC "qualifier" ion to the method. The role of the qualifier ion is to provide identity confirmation of the analyte being detected and it plays no part in the quantification process, for which the "quantifier" ion is used. Additionally, the cone voltage and collision energy were optimised for each compound. The former refers to the voltage applied to the guide cones that direct the ion beam from the ESI source to the mass spectrometer vacuum. The latter describes the electric potential applied the second quadrupole of the mass spectrometer instrument that leads to fragmentation of the

precursor ions. The example chromatograms obtained for all compounds analysed are shown in Supplementary Figure 1.

Table 3.3 The summary of mass spectrometry running conditions for each compound analysed

Compound name	Precursor ion mass to charge ratio	Product ion mass to charge ratio	Retention time (minutes)	Cone voltage (V)	Collision energy (V)
2′-deoxycytidine	228	112	5.67	40	12
5-methyl-2′-deoxycytidine	242	126	6.01	15	12
5-hydroxymethyl-2′-deoxycytidine (quantifier ion)	258	142	5.91	45	23
5-hydroxymethyl-2′-deoxycytidine (qualifier ion)	258	124	5.91	45	10
2'-deoxycytidine (^{15}N3)	231	115	5.67	40	12
5-methyl-2'-deoxycytidine-d3	245	129	6.01	15	12
5-hydroxymethyl-2'-deoxycytidine-d3 (quantifier ion)	261	145	5.91	45	23
5-hydroxymethyl-2'-deoxycytidine-d3 (qualifier ion)	261	145	5.91	45	10

3.4.2.2 Assessment of the potential sources of ion suppression in the sample preparation and liquid chromatography workflow

The main sources of possible ion suppression in the method were identified as DNA extraction, DNA degradation and HPLC conditions, including the column type, mobile phase flow rate and analyte dimerization. The DNA extraction process involves numerous incubation steps with buffers and enzymes that can potentially lead to contamination of the DNA with various salts and proteins. Commercially available DNA extraction and clean up kits were briefly considered, however a trial of a selected few of them failed to improve either the signal-to-noise ratio or the overall signal of the analysed compounds as compared to the traditional phenol:chloroform:isoamyl alcohol (PCI) protocol followed by the ethanol and sodium acetate precipitation (data not shown). These kits included the Quick-DNA Miniprep Plus Kit (Zymo Research) and KAPA Express extract kit (Kapa Biosystems) as well as the Genomic DNA Clean & Concentrator (Zymo Research) and Monarch PCR & DNA Cleanup Kit (New England Biolabs). The latter two kits were used in an attempt to clean up the DNA following the regular PCI extraction protocol. The DNA extraction method selected was hence the traditional PCI extraction.

It was still, however, entirely possible that the PCI method of DNA extraction contributed to ion suppression and hence caused an increase in the limit of detection for the analytes of interest. In order to test this hypothesis, 10 kDa Amicon filters (Sigma) used for desalting and protein removal were utilised. They were included to clean up the DNA samples isolated from the HCT116 colorectal cancer cell line or the breast cancer cell line MCF7 either post DNA extraction, post DNA digestion or both (Figure 3.6). Surprisingly, the signal-to-noise ratio was not significantly improved in any of the Amicon filter combinations in neither of the cell lines tested. This result, combined with the high cost of the columns, led to their exclusion from all future experiments.

Next, the DNA degradation process was optimised. The manufacturer of the DNA degradase enzyme suggests 2 hours as the optimal digestion time for obtaining single nucleotides from DNA molecules. It is, however, possible that this might not be a sufficient amount of time to allow for complete DNA degradation. In order to test this, a 24-hour DNA digestion time course was performed (Figure 3.7). The overall signal obtained for all three analytes was assessed with the expectation of a bell-shaped

response. Initially, an increased response would be observed as the digestion time was extended. This would be followed by a decrease in the signal due to the prolonged exposure of nucleotides to the high digestion temperature (37ºC). Indeed, the results indicate that approximately 8 hours of digestion appears to maximise the signal obtained on the mass spectrometer for the analytes of interest. In order to compromise between digestion efficiency and digestion time, however, 4 hours was chosen as the optimal digestion time for all future experiments.

The last stage of the optimisation process involved the manipulation of HPLC parameters in order to increase the overall signal obtained on the mass spectrometer, reduce the signal-to-noise ratio and improve the peak shape and resolution for the analytes. In order to achieve that, a HPLC column utilising a different chromatography method was compared to the C18 column used so far - namely the HILIC column (see section 3.2.1 for background on this method). It is also known that chemical additives can improve the chromatographic peak shape and resolution. Additionally, they can help with dissociating analyte dimers, which have been observed for all three nucleosides during instrument tuning (data not shown). In order to test this, 10 mM ammonium acetate was added to both mobile phase buffers while using the original C18 column. Neither the HILIC column nor the ammonium acetate managed to improve the peak shape or resolution and were hence not included in the future experiments (see Supplementary Figure 2 for comparison of chromatograms).

The flow rate of mobile phase through a HPLC column can affect the chromatographic peak shape and hence the signal obtained on the mass spectrometer. If the flow rate is set too high or too low, it can lead to peak broadening, peak splitting or shift to an earlier retention time. This can lead to less analyte binding to the column and a more gradual elution resulting in a reduced overall signal and lower sensitivity of the method. In order to establish the optimal flow rate for this method, a range of flow rates were tested - between 150 and 300 µL/min in increments of 50 µL/min (Supplementary Figure 3). By assessing the chromatograms obtained from this experiment, it is clear that the 150 µL/min flow rate is the optimal one for this method.

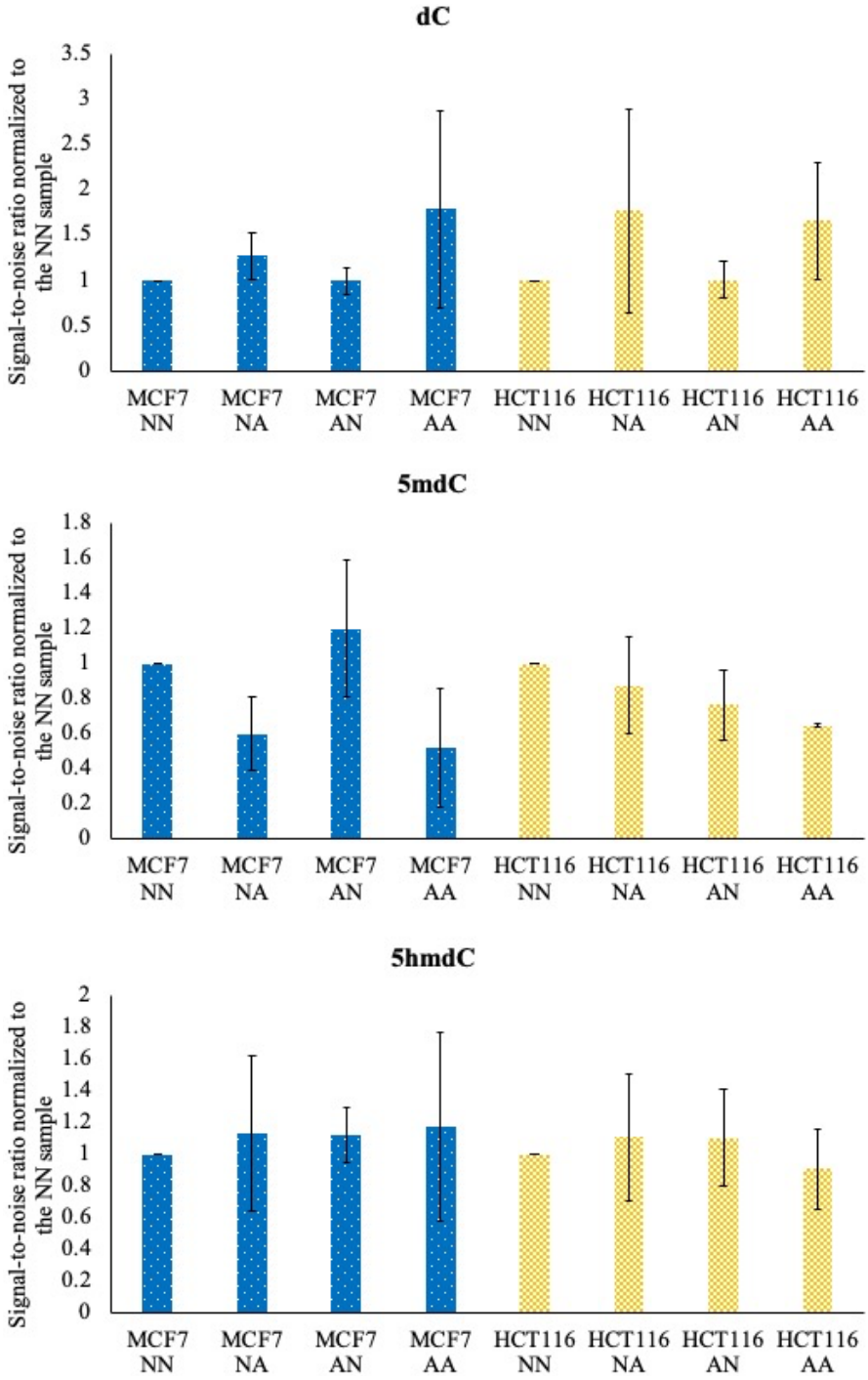


Figure 3.6 DNA clean up with Amicon filters has no effect on the signal-to-noise ratio in cell line samples

The following samples were analysed: NN - no Amicon filters were used; NA - Amicon filters were only used post DNA degradation; AN - Amicon filters were only used post DNA extraction; AA - Amicon filters were used both after the DNA extraction and DNA degradation. Three independent replicates were analysed for each sample. The error bars represent the standard deviation. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison.

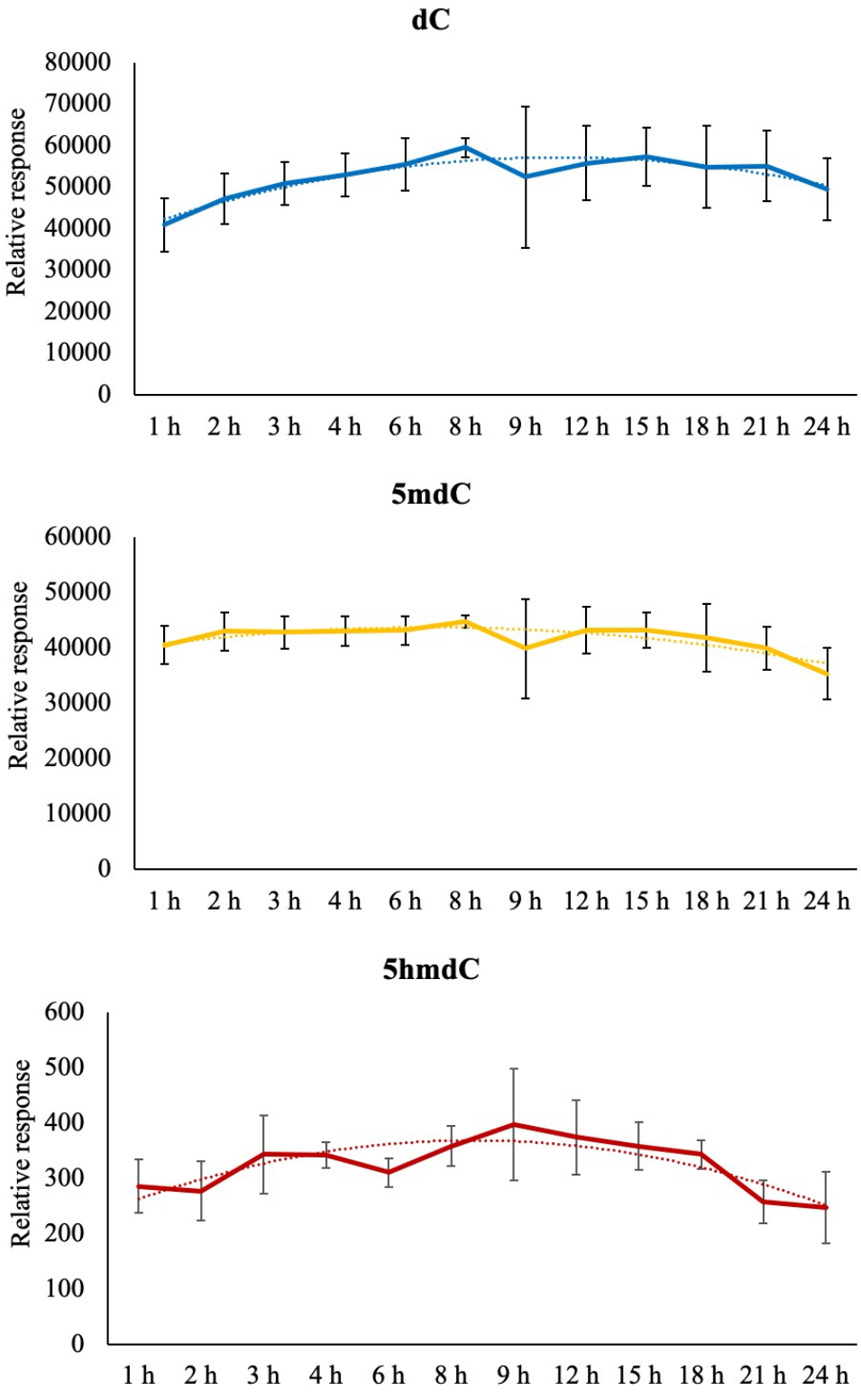


Figure 3.7 The highest signal for dC, 5mdC and 5hmdC is obtained after 8 hours of DNA digestion

The solid line represents a direct trend between the time points and the dotted line displays the line of best fit. The error bars represent the standard deviation of three independent replicates. The x-axis lists all the time points assessed in this experiment.

3.4.2.3 The method and instrument validation

The completion of this in-house mass spectrometry method optimisation was followed by assessment of the analyte losses during sample preparation (method validation) as well as the assessment of both intra-and interday variability of the mass spectrometer (instrument validation). The method validation experiment was carried out by introducing three known concentrations of synthetic standard mix into the samples either pre- or post-digestion to assess the analyte losses during the enzymatic reaction. The three concentrations are referred to as “High”, “Middle” and “Low”, as a commonly used system of quality control standards (QCs) in mass spectrometry to assess instrument performance. The exact values of these concentrations are based on the calibration curves prepared for each standard. These QCs are commonly used with each sample run to allow comparison of the signal it produces on different days and sample runs.

All three compounds showed a high degree of recovery during the digestion process (Figure 3.8). Only 5mC showed a significantly reduced recovery for medium and low QCs, however none of the relative signals obtained were lower than 80% of the post-degradation spikes. It was particularly encouraging to see no significant losses of 5hmC signal as it is the lowest abundance molecule of the three. This results also validates the previously chosen 4-hour digestion time as one that does not cause sufficient analyte degradation to cause issues with their detection.

The intraday instrument validation was performed using a similar set of QCs comprised of pure synthetic standards. The aim of the experiment was to inject the same concentration of the analytes three times in one day to assess the measurement variability inherent to the mass spectrometer. All samples were injected on the mass spectrometer in a technical triplicate and at different times of the day relative to each other, and hence it is critical to know what inherent variation between those injections should be expected because of experimental error.

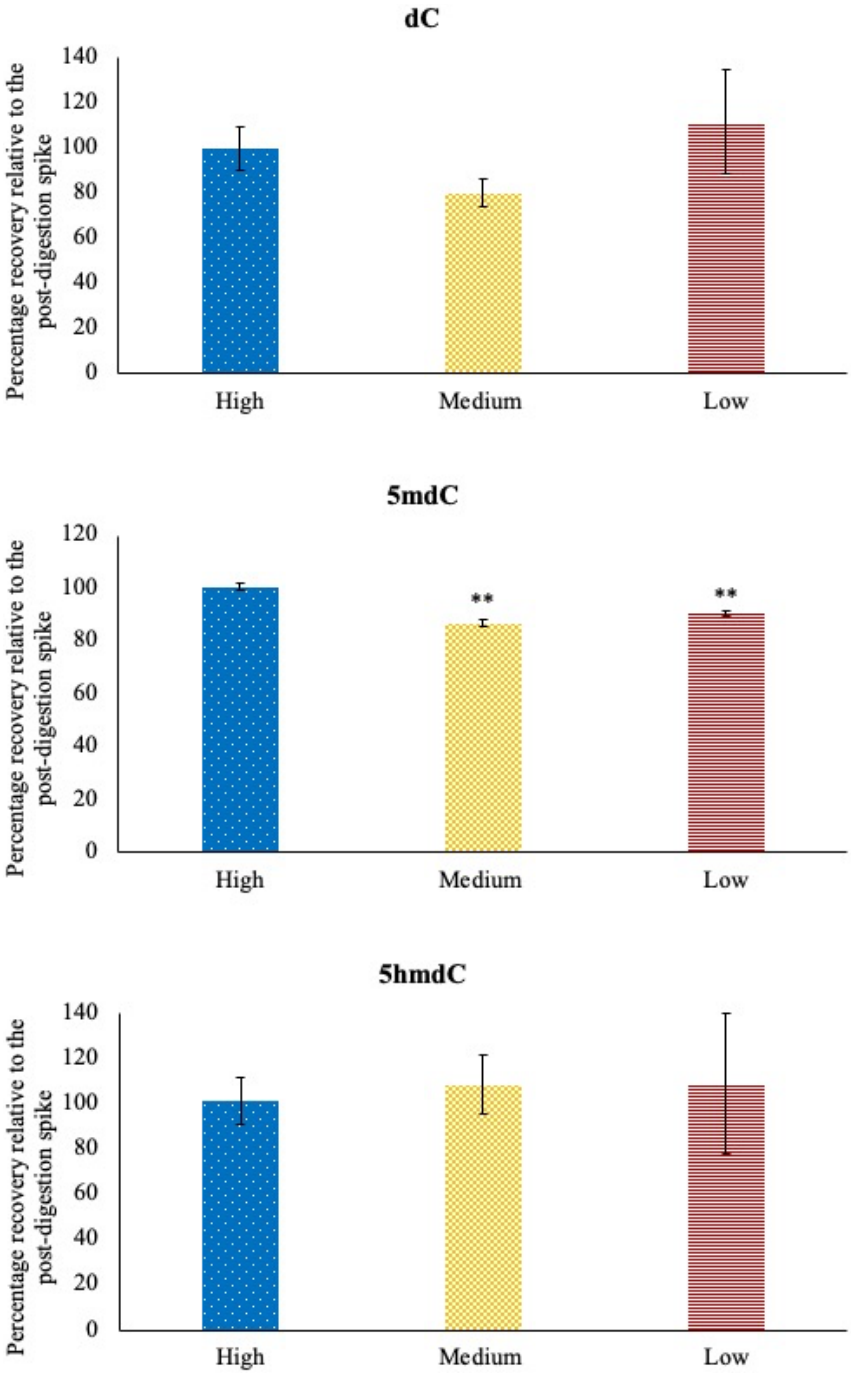


Figure 3.8 The three analytes show minimal losses during DNA digestion reaction
The SW620 cell line DNA was spiked with the High, Middle and Low QCs either pre- or post-digestion. The QCs used were as follows: 2300 nM (High, blue bar), 35.9 nM (Middle, yellow bar) and 8.98 nM (Low, red bar) for cytosine; 230 nM (High, blue bar), 3.59 nM (Middle, yellow bar) and 0.90 nM (Low, red bar) for 5mC and 5hmC. The results are shown as the percentage of the signal obtained for the pre-digestion spike samples divided by the signal produced by the post-digestion spike samples. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (** $p<0.01$).

The results outlined in Table 3.4 demonstrate that there is a degree of variation in the reproducibility of the mass spectrometry analysis that one can expect from the experimental set up itself. The percentage variation (a direct measure of the reproducibility) is below or around 10% for most of the synthetic standard samples tested. The only outlier is the Low QC cytosine sample. This is not of a particular practical concern, however, as one would never observe a cytosine concentration this low in the DNA samples. Interestingly, 5mC displays a considerably lower percentage deviation for all its QC samples compared to the other two molecules. This is in line with the general observation that 5mC has the lowest detection limit, the best shape of the HPLC peak and the highest degree of linearity of response across multiple calibration curves (data not shown).

Table 3.4 The summary of intraday instrument validation results

dC			
Sample	Mean relative response	Standard deviation	Percentage deviation
High QC (2760 nM)	0.3930	0.0322	8.20
Middle QC (690.0 nM)	0.0819	0.0090	11.0
Low QC (14.38 nM)	0.0011	0.0002	21.4
5mdC			
Sample	Mean relative response	Standard deviation	Percentage deviation
High QC (552 nM)	1.9708	0.0193	0.98
Middle QC (69.0 nM)	0.2315	0.0050	2.14
Low QC (1.43 nM)	0.0048	0.0001	1.15
5hmdC			
Sample	Mean relative response	Standard deviation	Percentage deviation
High QC (92.0 nM)	0.0261	0.00152	5.84
Middle QC (5.75 nM)	0.0016	0.00019	12.1
Low QC (0.36 nM)	0.0003	0.00003	11.1

The interday instrument validation was performed in a similar way to the intraday validation described above. The selected High, Middle and Low QC samples were injected on six different days to assess the reproducibility in the signal obtained on the mass spectrometer across independent experimental runs (Table 3.5). Prior to the

experiment, the hypothesis was that the variation observed will be considerably greater than the intraday one described above. While all the dilutions were prepared from the same aliquoted stock, the pipetting errors in a serial dilution series can compound to affect the signal obtained on the mass spectrometer on different days. In addition to that, the HPLC column performance can contribute to the variability.

Table 3.5 The summary of interday instrument validation results

dC			
Sample	Mean relative response	Standard deviation	Percentage deviation
High QC (3960 nM)	1.3003	0.5185	39.9
Middle QC (247.5 nM)	0.1122	0.0879	78.3
Low QC (3.867 nM)	0.0068	0.0066	97.5
5mdC			
Sample	Mean relative response	Standard deviation	Percentage deviation
High QC (396 nM)	1.7083	0.5725	33.5
Middle QC (24.8 nM)	0.1570	0.1118	71.2
Low QC (0.39 nM)	0.0077	0.0094	122
5hmdC			
Sample	Mean relative response	Standard deviation	Percentage deviation
High QC (396 nM)	0.1516	0.0535	35.3
Middle QC (24.8 nM)	0.0112	0.0041	36.6
Low QC (0.39 nM)	0.0005	0.0002	41.7

As predicted, the interday variability was indeed considerably higher for all compounds and samples as compared to the intraday values, ranging from 3-fold to as high as 106-fold increase in the percentage deviation values. Interestingly, the lowest variability was seen in the 5hmC samples, especially at the lowest concentration. This is indeed in stark contrast to the pattern described for the intraday variation results. In the light of these findings, it was decided to run all the samples that will be directly compared to each other on the same day, thus avoiding the poor interday reproducibility skewing the results.

3.4.3 The validation of the in-house protocol against a previously published external method

The last step of optimising the in-house mass spectrometry method was to compare its performance directly with an already established mass spectrometry method at an external facility. Alison Brewer's group at King's College London (KCL) previously successfully used their in-house mass spectrometry method to demonstrate the effect of oxygen gradient on 5hmC levels during hESC differentiation (Burr et al., 2018). The same method described in the paper was utilised in this case and was indeed capable of detecting 5mC and 5hmC in the hESC differentiation samples (Figure 3.9 A). It should be noted that the two mass spectrometry methods provided different absolute values for both 5mC and 5hmC levels in the same samples. The 5mC levels detected by the external method were between 8 and 9% of all cytosines compared to between 2.5 and 4% obtained with the in-house protocol. Similarly, there was also discrepancy in 5hmC levels measured: between 0.015 and 0.08% of all cytosines according to the external mass spectrometer as compared to between 0.1 and 0.4% seen with the in-house method. This is likely due to different calibration curves being used to normalise the results obtained on the two instruments and the inclusion of internal standards in the in-house analysis. Additionally, the results of the in-house experiment had a greater degree of variability in both the 5mC and 5hmC levels as shown by the larger error bars in Figure 3.9 B. This is likely an indication of lower intraday variability of the external method.

However, the external mass spectrometer failed to detect 5hmC in any of the colorectal cancer cell lines including the SW480 and SW620 cell lines characterised in the subsequent Chapter (Supplementary Figure 4). In contrast, the in-house mass spectrometry method described here was capable of robustly detecting 5hmC levels in all the cell line samples analysed indicating a higher sensitivity of the in-house protocol (Figure 3.9).

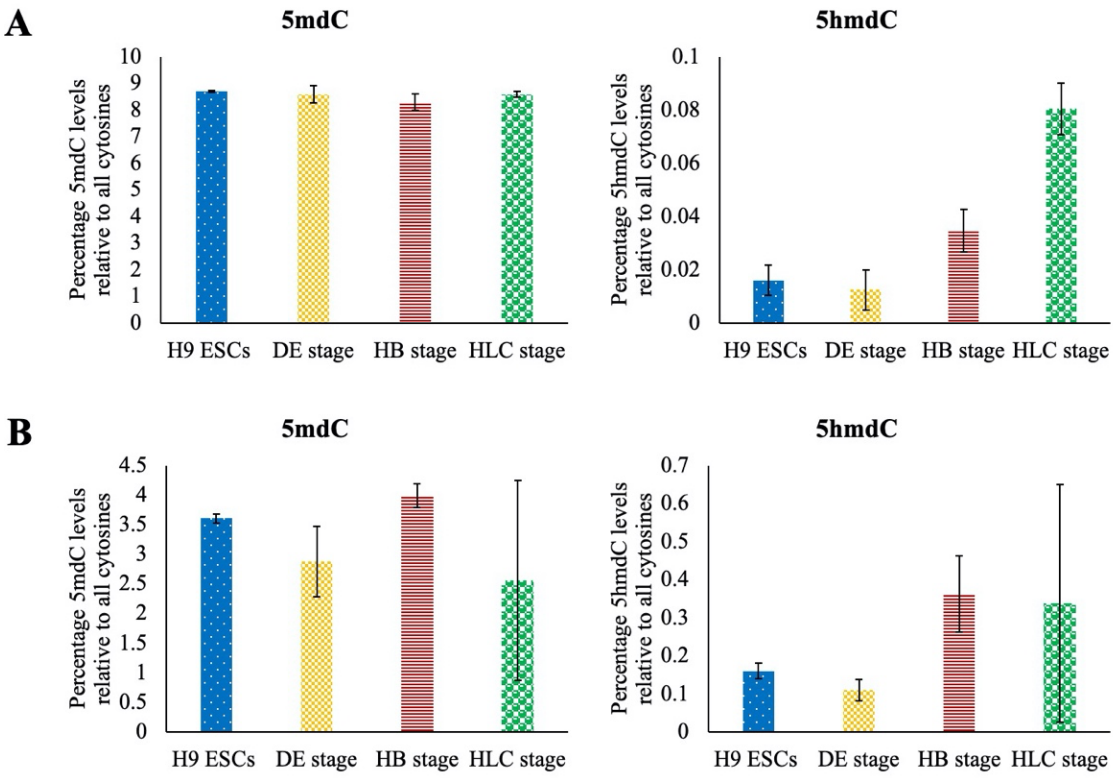


Figure 3.9 The similar pattern of changes of 5mC and 5hmC levels in the human embryonic stem cell differentiation samples using the in-house and external mass spectrometry methods

The 5mC and 5hmC levels measured in human embryonic stem cell differentiation samples by (A) the external KCL method; (B) the in-house mass spectrometry method. The following samples were analysed: H9 ESCs - human embryonic stem cell line (blue bars); DE stage - definitive endoderm stage of differentiation (yellow bars); HB stage - hepatoblast stage of differentiation (red bars); HLC stage - hepatocyte-like cell stage of differentiation (green bars). The error bars represent the standard deviation of two independent replicates analysed.

3.5 Discussion

In order to study the epigenetic changes in cancer progression and cellular differentiation, a sensitive quantification method of these processes must be employed. There are numerous methods of 5mC and 5hmC levels quantification on a global DNA scale. Dot blots offer a simple and inexpensive method of antibody-based 5mC and 5hmC detection (Jia et al., 2017). They are, however, only semi quantitative in nature and their limited sensitivity prevents them from being able to accurately show small changes in the levels of these nucleotides. They also require a considerable amount of starting DNA (usually a few micrograms) and are prone to errors due to uneven DNA loading or unspecific antibody binding.

Immunofluorescence and immunohistochemistry are also antibody-based and frequently used to visualise changes in levels of cytosine modifications across a tissue section (Kim et al., 2016; Chapel et al., 2019). They are a critical tool to show changes in 5mC and 5hmC between tissue types or during a process such as cellular differentiation of adult stem cells in the mouse small intestine (Uribe-Lewis et al., 2020). These methods are also only semi-quantitative and affected by the non-specific binding of antibodies. They cannot hence be used to accurately determine the changes in global 5mC and 5hmC levels. There are commercially available ELISA kits for 5hmC detection, such as the Quest 5-hmC DNA ELISA Kit (Zymo Research). They were found, however, to suffer from poor reproducibility and low sensitivity (Gilat et al., 2017). Lastly, thin layer chromatography was utilised prior to the widespread use of mass spectrometry in the field. While it was successfully used to detect 5mC, 5hmC and even the low abundance 5fC and 5caC, it is not itself quantitative and requires the use of radioactive isotopes (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). It is therefore clear why mass spectrometry became the gold standard in the field and why it was chosen as the preferred quantification method for this PhD project.

When developing a mass spectrometry-based method of detection of low-abundance analytes, it is critical to remove all potential sources of contamination. The most prominent of those are compounds and molecules that interfere with the process of ion formation following the chromatographic separation (the phenomenon known as ion suppression). There are numerous possible sources of ion suppression that can contribute to increased limit of detection and limit of quantification values. It was hence

critical to address these effects in a comprehensive way as part of the in-house method development given the low abundance of 5hmC in the genome. The choice of the correct mass spectrometer type for the analysis represents the logical first step of such development. To date, numerous mass spectrometer types have been used to assess the global 5hmC levels in DNA samples. These include triple quadrupole, QTOF and an orbitrap triple quadrupole instruments (Bachman et al., 2014; Thienpont et al., 2016; Guo et al., 2017). The triple quadrupole mass spectrometer was found to be the most suitable for this in-house analysis method with superior sensitivity over the QTOF instrument. Regrettably, the University of Bath does not own an orbitrap triple quadrupole instrument similar to the one used by Bachman and colleagues to allow comparison to the other two instrument types.

The Agilent 6490 triple quadrupole instrument has emerged as the gold standard instrument for detection of low-abundance cytosine modifications such as 5mC, 5fC and 5caC (Thienpont et al., 2016). The same instrument was also used by Petra Hajkova's group at Imperial College London to detect 5hmC levels in low quantities of DNA extracted from a small (as little as 100) number of cells (Amouroux et al., 2016; Hill et al., 2018). It is nearly impossible to directly compare the performance of this cutting-edge mass spectrometer to the Xevo TqD instrument used in the in-house method without running the same set of standards and samples on both machines. According to both manufacturers data, the Agilent instrument appears to be roughly 1000-fold more sensitive (Agilent, n.d.; Waters, 2019). It should be noted that different compounds were used for these measurements and hence this is just an approximation. The proprietary iFunnel Agilent technology likely contributes to the superior sensitivity of the instrument by providing an improved ESI and ion focusing that results in up to zeptomole detection limits (Agilent, n.d.).

A direct comparison can also be made between the in-house protocol and an external method. Bachman and colleagues have reported that the linearity of response in their orbitrap-based method extends to analyte quantities below 10 amol as compared to the in-house method which had linearity of response limit at 30 amol for 5mC and 120 amol for 5hmC (Bachman et al., 2014). Petra Hajkova's group published 5 amol as their limits of quantification (defined as signal to noise ratio above 10) for both 5mC and 5hmC (Amouroux et al., 2016). This is almost 400-fold lower than the in-house limits of quantification which were 1.9 fmol for both 5mC and 5hmC. These lower

quantification limits obtained by Hajkova's group reflect their need for superior sensitivity given the previously mentioned scarce amount of starting material. Nevertheless, it is important to recognise that low detection levels were obtained for the analytes in the in-house protocol while using a considerably less sensitive mass spectrometry instrument.

The DNA extraction and digestion processes have been identified as another potential source of ion suppression. In particular, the cell lysis and DNA digestion buffers contain multiple salts that may reduce ionisation efficiency. Numerous methods of sample clean-up have been used in the literature, including filtration plates or spin columns (Shen and Zhang, 2012; Thienpont et al., 2016; Mulholland et al., 2018; Ambati et al., 2019). In particular, the use of Amicon filters was previously published as a way of removing contaminants that could affect the ionisation process (Bachman et al., 2014). However, when tested, the extra filtration step failed to improve the signal-to-noise ratio that directly measures the level of sample contamination relative to the signal produced. This could potentially be due to the samples already having low level of contamination. Extensive ethanol precipitation, careful washes and substantial dilution of the sample following digestion could all be contributing to that. Similarly, various approaches to DNA digestion have been described previously. These ranged from a simple DNA degradase-based digestion to more complicated procedures involving Nuclease S1 and antarctic phosphatase or benzonase and alkaline phosphatase (Bachman et al., 2014; Amouroux et al., 2016; Thienpont et al., 2016). While the method described here relied on the DNA degradase enzyme, it was found that an increased digestion time improved the signal obtained on the mass spectrometer. Instead of the 2 hours suggested by the manufacturer, between 4 and 8 hours was found to give a higher signal for all three compounds analysed. The most likely explanation is that the longer digestion time provides the enzyme with an opportunity to complete the DNA digestion. This is despite using the Degradase enzyme in excess, as per manufacturer's instructions.

The method validation experiment provided an insight into the analyte losses during the digestion process. Prior to the analysis, it was hypothesised that the exposure of analytes to the digestion temperature of 37ºC and the buffer matrix of the digestion reaction could result in losses of said analytes. In particular, the added moieties on 5mC and 5hmC were thought to be prone to degradation, possibly through hydrolysis. It was indeed found that a few of the QC-spiked samples did lose a part of the analyte signal;

however, the majority of the samples had no significant losses. This finding provided additional evidence for the method being well optimised already for maximising the signal obtained on the mass spectrometer.

The investigation of intra- and interday variability offered an insight into the degree of reproducibility of the future experimental results. The intraday values were found to be relatively low - below or around 10% deviation for most QC samples. These values reflect the reproducibility of performance of the mass spectrometer and possible analyte degradation over time. Cytosine stood out as the analyte with the highest percentage deviation across the range of QC concentrations. This is in line with the compound having the broadest peaks during liquid chromatography, the highest detection limit and generally the lowest linearity across multiple calibration curves (data not shown). The high QC concentration, however, showed the lowest variability which is encouraging given that it is the concentration of cytosine most closely resembling the one observed in DNA samples. The very high reproducibility of the 5mC signal across all QCs was in line with its aforementioned high linearity and low detection limit in numerous calibration curves. Lastly, 5hmC signal showed a slightly higher degree of deviation from the mean. It should be noted, however, that all 5hmC QC concentrations are considerably lower than the equivalent QCs for the other two compounds which could be contributing to the variation.

The interday variability was predicted to be considerably higher than the intraday counterpart prior to the experiment. This is due to a drastically increased number of factors that can contribute to the changes in results obtained on different days. These include, among others, liquid chromatography column and mass spectrometer performance, pipetting error during calibration curve dilutions, time taken before the QC preparation and running as well as inherent variation in the analyte concentration in the frozen master stocks from which the dilutions are prepared. Indeed, the interday reproducibility was found to be considerably lower than the intraday one for all analytes and QC concentrations assayed. In particular, the values were surprisingly high for 5mC, with the low QC concentration showing over 100% deviation from the mean. This high variation, while concerning, was not deemed particularly problematic for future analysis as it was decided to run all samples that will be directly compared to each other on the same day. This *de facto* bypasses the issue of poor interday reproducibility. Additionally, for the vast majority of samples they are compared to one another in terms

of their 5mC and 5hmC levels rather than having the absolute levels of these molecules established in them independently. As such, as long as each independent replicate set of samples is ran on the same day, even when the three independent replicates are ran on different days, the interday variation will not affect the results.

When developing a mass spectrometry method in-house, it is critical to compare its performance to a previously established one in order to gain a better understanding of its performance and limitations. In this case, the in-house method was compared to a previously published protocol from Dr Brewer's group at KCL. Both methods displayed similar performance when assessing the 5mC and 5hmC levels in samples from a protocol of human embryonic stem cell differentiation to hepatocyte-like cells (see Chapter 6 for a detailed description of the process). This was to be expected given high levels of these modifications previously found in the H9 ESCs and a similar hESC differentiation experiment to a pancreatic lineage (Li et al., 2018; Koutsouraki et al., 2019). The external mass spectrometer was not, however, able to detect 5hmC in a number of colorectal cancer cell lines. These cell lines are known to have low 5hmC levels, possibly due to their high proliferation rate (Li and Liu, 2011; Bachman et al., 2014; Uribe-Lewis et al., 2015; Huang et al., 2016). In stark contrasts, however, 5hmC was readily detectable in the same cell line samples on the in-house triple quadrupole mass spectrometer. This is of particular importance given that the subsequent Chapters mostly describe work done in cell line models. This finding hence validated the in-house method as one that is more sensitive than a previously published external method used by Dr Brewer's group at KCL.

There are still, however, a number of steps that could be taken to further improve the method described here. The first clear limitation of this protocol is the inferior detection compared to the previously published methods described above. As alluded to earlier, this might be caused by the lower sensitivity of the mass spectrometer instrument itself. One could, however, make further attempts at eliminating potential sources of ion suppression. Following the extraction process, the DNA could be vacuum- or nitrogen gas-dried prior to reconstitution in LC-MS grade water to remove the traces of ethanol prior to digestion (Guo et al., 2017; Ambati et al., 2019). Similarly, the samples could be filtered following DNA degradation to remove the degradase enzyme and other protein contaminants, thus potentially improving the chromatographic separation and detection sensitivity (Thienpont et al., 2016). Further testing of different C18 columns

of varying lengths and pore sizes could also have a beneficial effect on the peak shape and width of the analytes of interest. The second major limitation of the in-house mass spectrometry method is the poor interday reproducibility for all three analytes of interest. This could potentially be addressed by assessing the effect of the length of exposure of the synthetic standard calibration curves to room temperature on the stability of cytosine and its derivatives. Similarly, large stocks of the three QC standard mixes could be prepared and stored to be used with every run, in order to minimise the error introduced by the diluting process. It is likely, however, that some of the interday variation observed is due to the differences in performance of both liquid chromatography and mass spectrometry and it is highly unlikely that one would be able to eliminate it completely.

The reduction of global 5hmC levels in cancer cells compared to tumour tissue makes it an attractive potential diagnostic biomarker (Chen et al., 2013; Li et al., 2017; Tong et al., 2019). Mass spectrometry, given its quantitative nature, reliability and sensitivity could be the tool of choice for analysis of patient samples for diagnostic purposes. While the major limiting factor for the implementation of this method is the cost of the instrument itself (Chowdhury et al., 2017), there is a possibility of collaboration between hospitals and academic institutions that already own suitable mass spectrometers. Furthermore, it was recently shown that 5hmC levels were negatively correlated with the metastatic potential of papillary thyroid carcinomas (Tong et al., 2019). This hints at the possibility of tools such as the method described in this chapter to be used for predicting the potential of tumours to become metastatic, although methods relying on assessment of 5hmC status at specific, previously validated genomic regions are more likely to be widely adopted in the clinical practice in the future (see Chapter 7 for more details). It is important to note, however, that such a use of this (and other previously established in the literature) mass spectrometry methods would have to be thoroughly validated first and at this point this hypothesis is purely speculative.

Chapter 4

Modulation of 5-hydroxymethylcytosine levels in an *in vitro* colorectal cancer metastasis model

4.1 Chapter summary

The aim of this study was to characterise the *in vitro* colorectal cancer metastasis model comprising of the primary SW480 and metastatic SW620 cell lines isolated from the same patient and to assess the impact of compound-based modulation of global 5hmC levels on the proliferation and migration of these cells. This work was undertaken based on the preliminary evidence showing the possible role of the TET enzymes and 5hmC in colorectal cancer metastasis.

The primary SW480 cell line was found to have elevated 5mC and 5hmC levels as well as *TET2* expression compared to its metastatic SW620 counterpart. *TET1* levels have the opposite expression pattern, showing an increase in the SW620 cells while *TET3* expression is unchanged between the cell lines. The TET2 and TET3 protein levels were found to follow the same pattern as transcript expression data while the TET1 content was found to be the same in the two cell lines. The SW480 cell line displays an increased migration rate compared to SW620 cells while the metastatic cell line shows a higher proliferation rate and colony formation than the primary one.

Vitamin C treatment of the SW480 and SW620 cells resulted in the expected increase in 5hmC levels but had no effect on the expression of the TET enzymes. No changes in either were found with the dimethyl 2-oxoglutarate and nickel (II) chloride treatments, despite the use of previously published doses. Low dose (0.1 mM) vitamin C was found to have no effect on cellular migration in the two cell lines. The ascorbic acid-induced reduction in proliferation and colony formation was rescued with the addition of catalase, indicating that it was caused by hydrogen peroxide-induced cytotoxicity rather than 5hmC-mediated effects.

Taken together, these results provide a detailed characterisation of the SW480 and SW620 cell lines as well as the in-depth description of the effects of ascorbic acid treatment of these cells.

4.2 Introduction

The development of cancer is a highly complex process that involves a cascade of events and interactions with the surrounding environment. Aside from the genetic, epigenetic and metabolic intracellular changes that drive cancer progression, there is a wide range of external factors that contribute to tumorigenesis. These include, but are not limited to, interaction with the immune system and the local stroma as well as nutrient and oxygen availability dictated by the presence and distribution of blood vessels. A hallmark of the cancer cells is their ability to migrate and invade the surrounding tissue, leading to the process of metastasis (Hanahan and Weinberg, 2000). The role of epigenetics, and more specifically 5hmC, in metastasis still remains poorly understood (see section 1.5.4 for a detailed description). The Murrell group has preliminary data obtained from CRC patients suggesting a possible role of 5hmC and the TET enzymes in the metastatic process (Figure 4.1). Global 5hmC levels were reduced in the primary adenocarcinomas compared to the normal colon tissue (Figure 4.1 A). However, the global 5hmC levels were elevated in the liver metastasis biopsies as compared to the primary cancer. While it is possible that the metastatic niche in the liver itself causes the cancer cells to increase their 5hmC content, an alternative interpretation of this result is that the elevation in 5hmC levels could be a driver of the metastatic cascade, possibly by activating genes associated with motility and invasion.

Additionally, data from independent expression databases shows that high expression of *TET2* and *TET3* transcripts leads to a shorter time to metastasis development while high expression of *TET1*, *TET2* and *TET3* transcripts results in shorter survival as compared to low expression of these enzymes in CRC patients (Figure 4.1 B). Taken together, this data suggests that the investigation of the role of TET enzymes and 5hmC in CRC metastasis warrants further investigation.

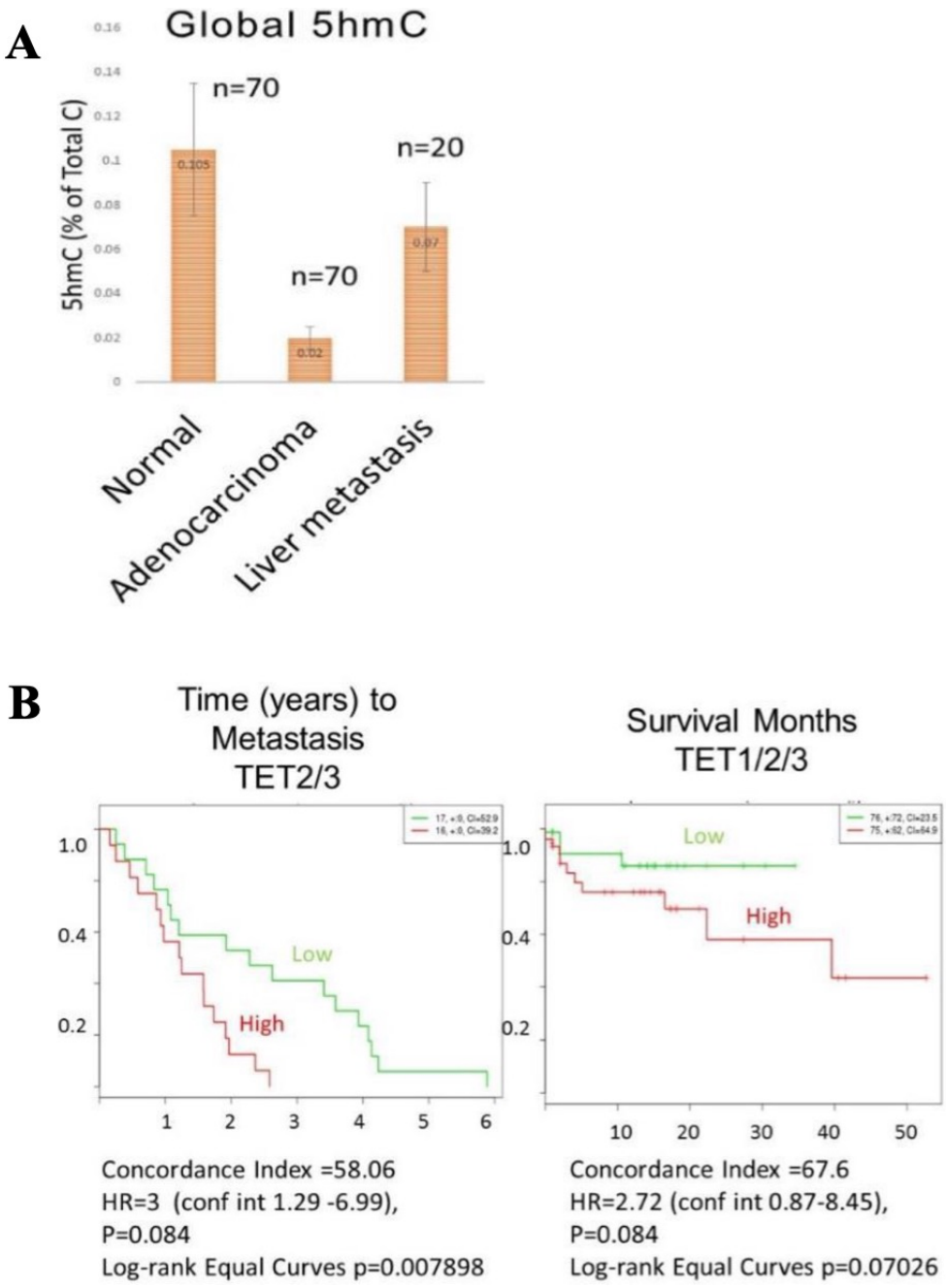


Figure 4.1 Preliminary data supporting the potential role of 5hmC and the TET enzymes in colorectal cancer metastasis

(A) The 5hmC levels measured in patient samples from a normal colon, adenocarcinoma of the colon and liver metastases by mass spectrometry. This unpublished result was obtained as a follow-up to the study on the role of 5hmC in colorectal cancer development (Uribe-Lewis et al., 2015).

(B) Survival curves generated using the PROGgeneV2 or SurvExpress tool and colon cancer expression databases (GSE28814 and GSE28722) indicate that that 'low-*TET2/3*' patients have a better metastasis-free survival. In a different study (GSE17536) with a similar number of patients (n>170), the overall 3-year survival rate for patients with high *TET1/2/*3 is lower than those with low *TET1/2/3* (Murrell, unpublished data).

4.2.1 *In vitro* modelling of colorectal cancer metastasis - the SW480 and SW620 cell lines

When studying the processes of tumour progression and metastasis, it is desirable to mimic these conditions as closely as possible via the use of *in vivo* models. Murine models remain the most widely used tool for studying tumour progression, metastasis and promising therapeutic interventions (Khanna and Hunter, 2009; Cekanova and Rathore, 2014). Murine models can be further subdivided into chemically induced, genetically engineered or transplantation ones. While all of these have their advantages, work with mice requires dedicated facilities, complex ethical approvals and trained staff. It is also more costly and less time efficient than the *in vitro* experimental techniques (Cekanova and Rathore, 2014).

The *in vitro* cancer models, initially in the form of immortalised cell lines, have been utilised since the early second half of the 20th century (Mirabelli et al., 2019). The simplest form of these models is a cell line grown in two-dimensional culture. This was improved by introducing a co-culture system where the selected cell line is cultured in the presence of stromal cells, most notably fibroblasts (Mirabelli et al., 2019). This allows one to account for the effects of the interactions with cells in the tumour microenvironment on cancer cells' response to various treatments or physical conditions.

The model chosen to investigate the effect of altering TET enzymes expression and modulating the 5hmC levels included two CRC cell lines: SW480 ([SW-480] ATCC CCL-228) and SW620 ([SW-620] ATCC CCL-227), which were obtained from the same patient (Leibovitz et al., 1976). The SW480 cell line was isolated from a 50 year-old Caucasian male's adenocarcinoma of the colon. These cells are cuboidal in shape and form small colonies of cells with epithelial-like morphology. Leibovitz and colleagues also observed that the cells have small villi projections on their surface (Leibovitz et al., 1976). The cells were found to be capable of forming tumours in nude mice with low rate of metastasis to the lung also reported (Tomita et al., 1992; De Both et al., 1999; Ma et al., 2008). The number of chromosomes per cell was estimated at 55 with the cells belonging to B type (the tumour has invaded the muscle surrounding the lining of the colon) on Duke's tumour classification based on the original biopsy taken from the patient.

The SW620 cell line was isolated from a lymph node metastasis from the same patient as the SW480 line (Leibovitz et al., 1976). It was cultured from a lymph node biopsy a year after the SW480 cell line was isolated from the primary tumour site. This was due to the cancer recurrence in the patient with a widespread metastasis throughout the abdomen. These were classified as Dukes type C adenocarcinoma (indicating that the tumour has spread to the neighbouring lymph nodes). The cells have a mostly smooth cell surface under the electron microscope and a spherical morphology with a number of elongated cells also present in the culture (Leibovitz et al., 1976). The number of chromosomes was estimated at 54. The SW620 cells were found to be highly tumorigenic in mice, forming metastases in lymph nodes and liver (Céspedes et al., 2007; Vuletic et al., 2017; Xu et al., 2020).

These two cell lines hence form a limited and simplistic model of CRC metastasis, which nonetheless could be useful for preliminary *in vitro* studies of the process. These cells also exhibit changes in the transcript levels of the three TET enzymes. This was revealed through an initial assessment of TET enzymes mRNA levels via qRT-PCR in a panel of twelve CRC cell lines (Supplementary Figure 5) performed in collaboration with Professor Ann Williams at the University of Bristol. The results demonstrated an over three-fold upregulation of *TET1* mRNA levels and a three-fold downregulation of *TET2* in the SW620 cells compared to the SW480 cell line, while the levels of *TET3* were similar in the two cell lines. Based on the origin of the two cell lines and the changes in *TET* expression described above, these were selected for further assessment during this PhD project.

4.2.2 Assessment of proliferation and migration in cell lines derived from primary and metastatic tumours

The SW480 and SW620 cell lines have previously been studied as *in vitro* models of metastasis, especially in terms of their proliferation and migration. Zhang and colleagues reported that the SW620 cell line exhibits a higher proliferation rate *in vitro* that the SW480 cells (Zhang et al., 2017). These results were in contradiction to two other publications whose authors found no changes in the proliferation rate between the two cell lines (Dai et al., 2017; Chen et al., 2019). The *in vitro* migration of SW480

cells was found to be higher than that of SW620 cells (Dai et al., 2017; Zhang et al., 2017). This was also true when *in vitro* invasion of the two cell lines was compared (Chen et al., 2019). A study of mouse xenografts found a significantly higher growth rate of the SW620 tumours *in vivo*, supporting the above mentioned *in vitro* results obtained by Zhang and colleagues (Hewitt et al., 2000; Zhang et al., 2017). The same study found that while the SW480 cells display a higher migration rate *in vitro*, the SW620 cell line showed an increased invasiveness *in vivo* as measured by a greater number of metastatic lesions formed in the animal (Hewitt et al., 2000).

4.2.3 Compound-based modulation of TET activity

The TET enzymes belong to a group of iron (II) and α-ketoglutarate dependent dioxygenase enzymes. Their activity depends not only on post-translational protein modifications such as acetylation (Zhang et al., 2017), phosphorylation (Bauer et al., 2015; Zhang et al., 2019), PARylation (Ciccarone et al., 2015) and *O*-GlcNAcylation (Shi et al., 2013) but also on the availability of oxygen, iron (II) and α-ketoglutarate. The levels of 5hmC in the DNA of a cell should therefore be modulated by compounds or physiological conditions that affect the abundance of either of these three factors. The most extensively studied to date compound that is capable of that is ascorbic acid (AA), also known as vitamin C. The Nobel Prize winner Linus Pauling proposed that vitamin C has efficacy in cancer treatment (Cameron and Pauling, 1976). However, the clinical trials conducted at the Mayo Clinic in the United States in late 1970s disproved these claims (Creagan et al., 1979). More recently, evidence emerged indicating that vitamin C as an adjuvant to chemotherapy has beneficial effects on the immune system and potential protective effect against the toxicity of the treatment (Ohno et al., 2009; Klimant et al., 2018; Pawlowska et al., 2019). Additionally, cancer patients are known to suffer from vitamin C deficiency and hence the benefits of AA supplementation may stem from there as well (Carr and Cook, 2018; Cimmino et al., 2018).

Vitamin C was also shown to act as a co-factor for the TET enzymes and enhance generation of 5hmC (Minor et al., 2013). Since then, it has been demonstrated to contribute to pluripotency maintenance and cell reprogramming of iPSCs (Stadtfeld et al., 2012; Blaschke et al., 2013; Gao et al., 2013; Schwarz et al., 2014). It was shown to

inhibit cancer hallmarks, including unchecked growth and migration, in numerous different cancer types *in vitro* (Gustafson et al., 2015; Ge et al., 2018; Peng et al., 2018; Sant et al., 2018; Shenoy et al., 2019). Interestingly, more recently it was demonstrated that vitamin C acts as a reducing agent in conversion of iron (III) to iron (II) rather than a co-factor for the TET enzymes as was previously thought (Hore et al., 2016). This provided a direct mechanism of action of AA to produce an upregulation in global 5hmC levels and highlighted the importance of iron (II) level for epigenetic regulation of the DNA. This role of vitamin C is particularly important given that iron (III) is the predominant intracellular form of this element (Hore et al., 2016).

The TET activity also depends on the levels of intracellular metabolite α-ketoglutarate, also known as 2-oxoglutarate. The Krebs cycle is responsible for α-ketoglutarate generation through the action of IDH1 and IDH2 enzymes (Chou et al., 2016). This molecule therefore acts as an intermediate between intracellular metabolic state and the epigenetic machinery. Higher intracellular levels of 2-oxoglutarate were shown to elevate the 5hmC content of the DNA by promoting 5mC oxidation by the TET enzymes (Letouzé et al., 2013). Despite that, it has not been used in research of cancer epigenetics to the same extent as AA. While cell-permeable dimethyl α-ketoglutarate has been utilised on numerous occasions in metabolic research (Zhang et al., 2014; Carey et al., 2015; Bott et al., 2019), it has not been extensively researched in terms of its effect on the cytosine methylation dynamics in cancer. In a model of pancreatic ductal adenocarcinoma, α-ketoglutarate was found to be a key mediator of p53 driven tumour suppression (Morris et al., 2019). At least a part of that effect was attributed to the resulting increase in 5hmC levels. It was additionally reported that 2-oxoglutarate may lead to an increased TET activity via a MYC-driven mechanism in B cell lymphoma (Qiu et al., 2020). Intriguingly, the authors also demonstrated that this pathway leads to an increased nuclear accumulation of the TET proteins and suggested that it could be contributing to tumorigenesis by demethylation of cancer-promoting promoters and enhancers. The influence of this metabolite on epigenetic profiles and properties of cancer cells warrants further investigation.

The activity of TET enzymes can also be inhibited by a number of compounds. Accumulation of metabolites such as succinate and fumarate has previously been demonstrated to lead to a reduction in TET activity (Xiao et al., 2012; Laukka et al., 2016). Similarly, 2-hydroxyglutarate produced by mutated IDH1 and IDH2 enzymes is

a competitive inhibitor of the TET enzymes, occupying the same site as α-ketoglutarate (Xu et al., 2011). Dimethyloxalylglycine, a synthetic inhibitor and a structural analogue of 2-oxoglutarate, was previously used to assess the effects of TET inhibition in multiple model systems (Zhao et al., 2014; Amouroux et al., 2016; Mahé et al., 2017). Lastly, nickel (II) ions were described as a potent inhibitor of 5hmC formation using a purified catalytic domain of TET1 as a model system (Yin et al., 2017). Due to structural similarities, nickel (II) was able to displace iron (II) from the enzyme's active site. A year later, the same group demonstrated that nickel (II) ions administered in the form of nickel chloride hexahydrate reduced 5hmC levels in human somatic cell line as well as mESC line (Yin et al., 2018).

There are numerous methods of compound-based modulation of global 5hmC levels in cells' DNA. Each of these has a number of off-target effects and affects a different part of the cell physiology. This chapter assesses how vitamin C, dimethyl 2-oxoglutarate and nickel (II) chloride alter 5hmC levels and their phenotypic effect on proliferation and migration in a primary (SW480) and metastatic (SW620) CRC cell lines.

4.3 Experimental objectives

Aim: to characterise the *in vitro* properties of the two colorectal cancer cell lines, SW480 and SW620, selected as a simplistic metastasis model and the effects of compound-based 5hmC levels modulation on the *in vitro* migration and proliferation rates of these cells.

Objectives:

1. Establish the TET expression, TET protein content and 5hmC levels in SW480 and SW620 cell lines
2. Establish the baseline parameters of proliferation and migration in the SW480 and SW620 cells
3. Assess the impact of vitamin C, dimethyl 2-oxoglutarate and nickel (II) chloride on TET expression and abundance of 5hmC in SW480 and SW620 cell lines
4. Assess the impact of vitamin C on proliferation, colony formation and migration rates of SW480 and SW620 cells

4.4 Results

4.4.1 Establishment of the TET expression, TET protein content and 5hmC levels in SW480 and SW620 cell lines

The SW480 and SW620 CRC cell lines were chosen for this investigation because they originate from the same patient at two different stages of cancer progression. Since the Murrell group has preliminary data showing that 5hmC levels are increased in metastatic liver tissue relative to the primary tumour in a subset of CRC patients (Figure 4.1 A), the levels of *TET* transcription were examined by qRT-PCR (Figure 4.2). *TET1* was significantly upregulated (4.6-fold) and *TET2* significantly downregulated (approximately 9-fold) in the metastatic SW620 cell line as compared to the primary CRC SW480 cell line (Figure 4.2 A). The transcript levels of *TET3* were unchanged. It should be noted that *ACTB* was selected as a reference gene using the GeNORM technique (see section 2.6.4 for details) which allows one to select the most stably expressed reference genes in a given system (Vandesompele et al., 2002).

The same pattern of expression was observed when the absolute mRNA levels (Figure 4.2 B) of each TET enzyme were determined through the use of standard curves as described in section 2.7. HCT116 CRC cell line, which was previously extensively studied by the Murrell research group and shown to have similar TET levels as CRC patient samples, was also included in an analysis as a control sample (Uribe-Lewis et al., 2015). The pattern of *TET* transcripts expression in SW480 and SW620 closely resembled that obtained through the relative expression measurements shown in Figure 4.2 A. *TET3* had the highest expression of all the *TETs* in both HCT116 and SW620 cell lines (0.5 and 0.9 pg of vector, respectively) while *TET2* was the most abundant of the three transcripts in the SW480 cell line (1.4 pg). *TET1* was found to have the lowest expression of all three *TETs* in both HCT116 and SW480 cell lines (0.1 pg), while in the SW620 cell line *TET1* expression was found to be higher than that of *TET2* (0.3 and 0.2 pg, respectively), the lowest abundance *TET* transcript in that cell line. The total levels of all *TET* transcripts were lower in the HCT116 cell line (0.8 pg) than both in SW480 and SW620 cells (2.5 and 1.4 pg, respectively). This likely reflects the difference in origin of these cell lines.

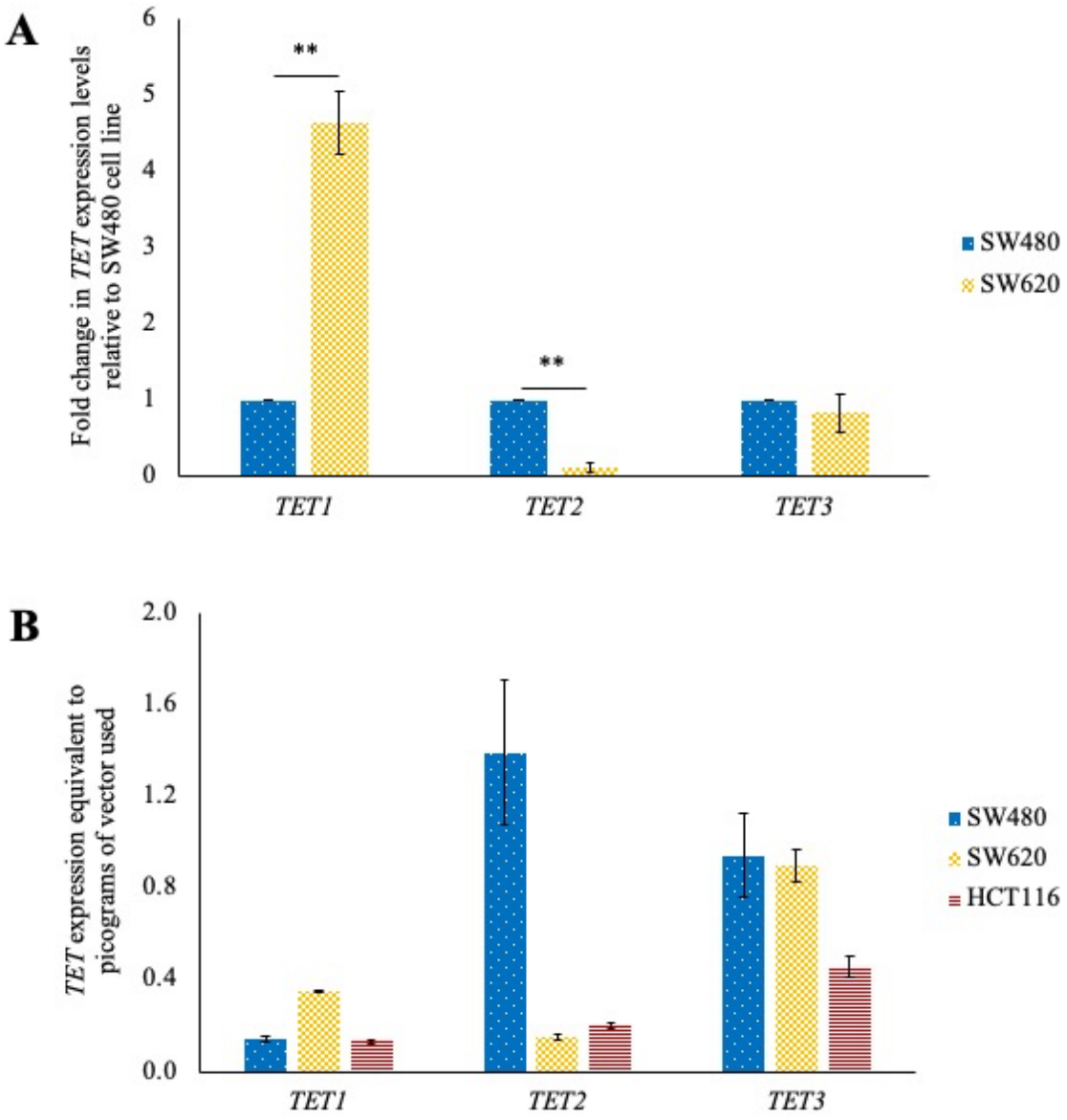


Figure 4.2 *TET1* and *TET2* transcripts display different expression patterns in SW480 and SW620 cell lines

(A) The expression of the *TET* transcripts in SW620 cell line relative to SW480 cell line. The *TET* mRNA levels were normalised to *ACTB* reference gene. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (** $p<0.01$).

(B) The absolute expression levels of the *TET* transcripts in HCT116, SW480 and SW620 cell lines equivalent to the picograms of plasmid used for preparation of a serial dilution standard curve. The error bars represent the standard deviation of three independent replicates.

The levels of the three TET proteins in the SW480 and SW620 cell lines were assessed using the Western blotting technique. The representative image of the antibody staining is shown in Figure 4.3 A. The full images of the blots for each of the TETs demonstrate the presence of multiple non-specific bands following staining with anti-TET1 and anti-TET3 antibodies. The amount of TET2 protein was found to be significantly (3.9-fold) higher in the primary SW480 cell line compared to SW620 cells (Figure 4.3 B). Surprisingly, no significant difference in TET1 levels was found. In agreement with the qRT-PCR results, the TET3 content was observed to be the same in the two cell lines.

Following the assessment of the mRNA and protein levels of the TET enzymes in the CRC cell lines, the abundance of 5mC and 5hmC in their DNA was measured (Figure 4.4) using both dot blot and mass spectrometry techniques. Dot blot analysis detected no discernible differences between these cell lines in terms of their 5mC and 5hmC content (Figure 4.4 A) whereas mass spectrometry analysis showed that the levels of 5mC were significantly reduced in the metastatic SW620 cell line compared to the two cell lines isolated from primary colorectal carcinomas - SW480 and HCT116 (Figure 4.4 B). The global levels of 5hmC was also significantly lower in SW620 cells compared to SW480 cells. The SW620 cell line had 1.6% 5mC and 0.03% 5hmC (out of all cytosines) levels, while 5mC values were 3.7% and 3.1% and 5hmC content 0.10% and 0.11% for HCT116 and SW480 cells, respectively. The dot blot technique is known to be only semi-quantitative, whereas mass spectrometry is more sensitive and quantitative over five orders of magnitude for 5hmC quantification as described in section 3.4.1. The reduction in 5hmC levels in the metastatic cell line is contrary to what was expected given the increase in *TET1* expression and the unpublished observation of increased 5hmC levels in liver metastases of a subset of CRC patients.

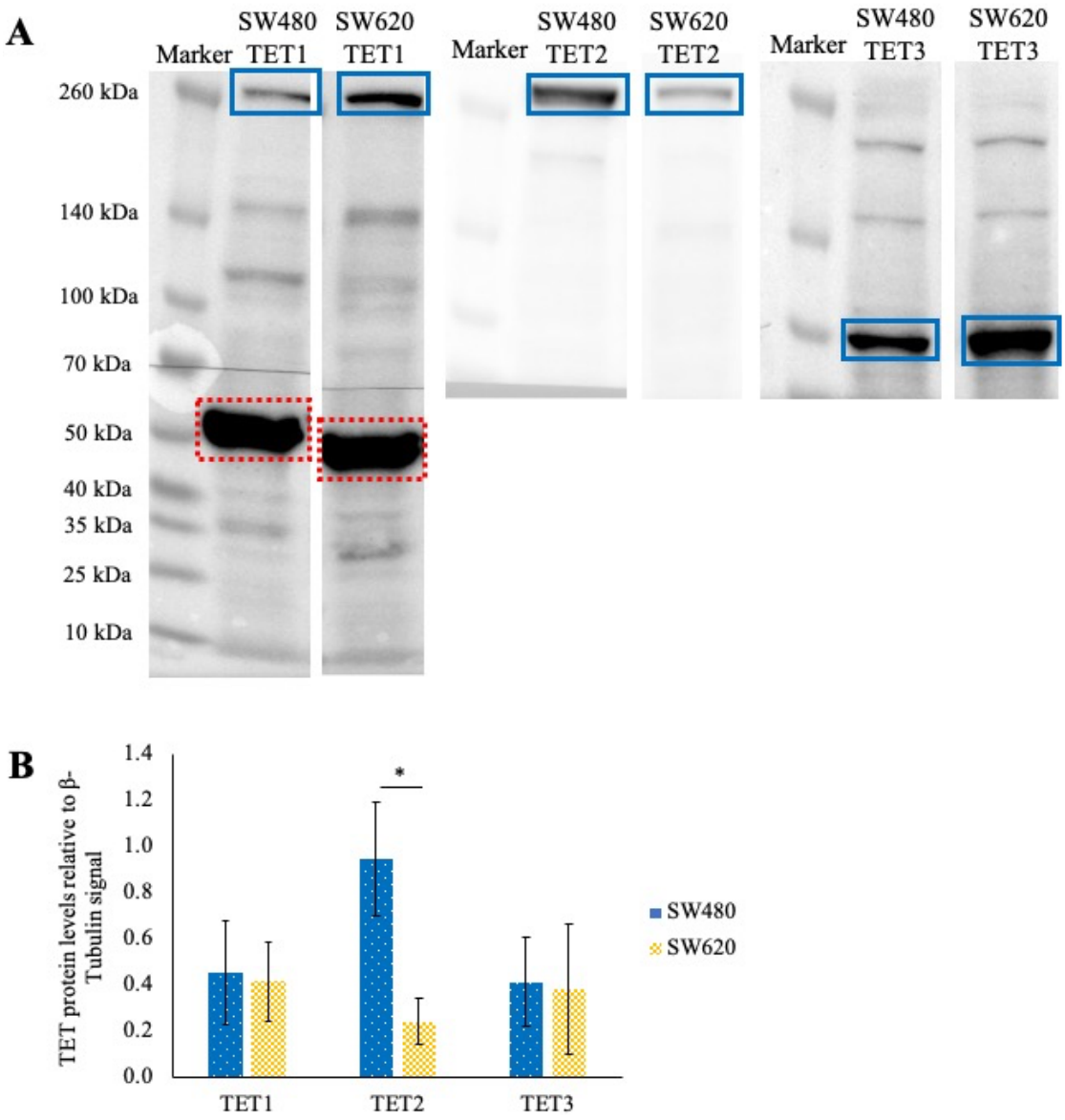
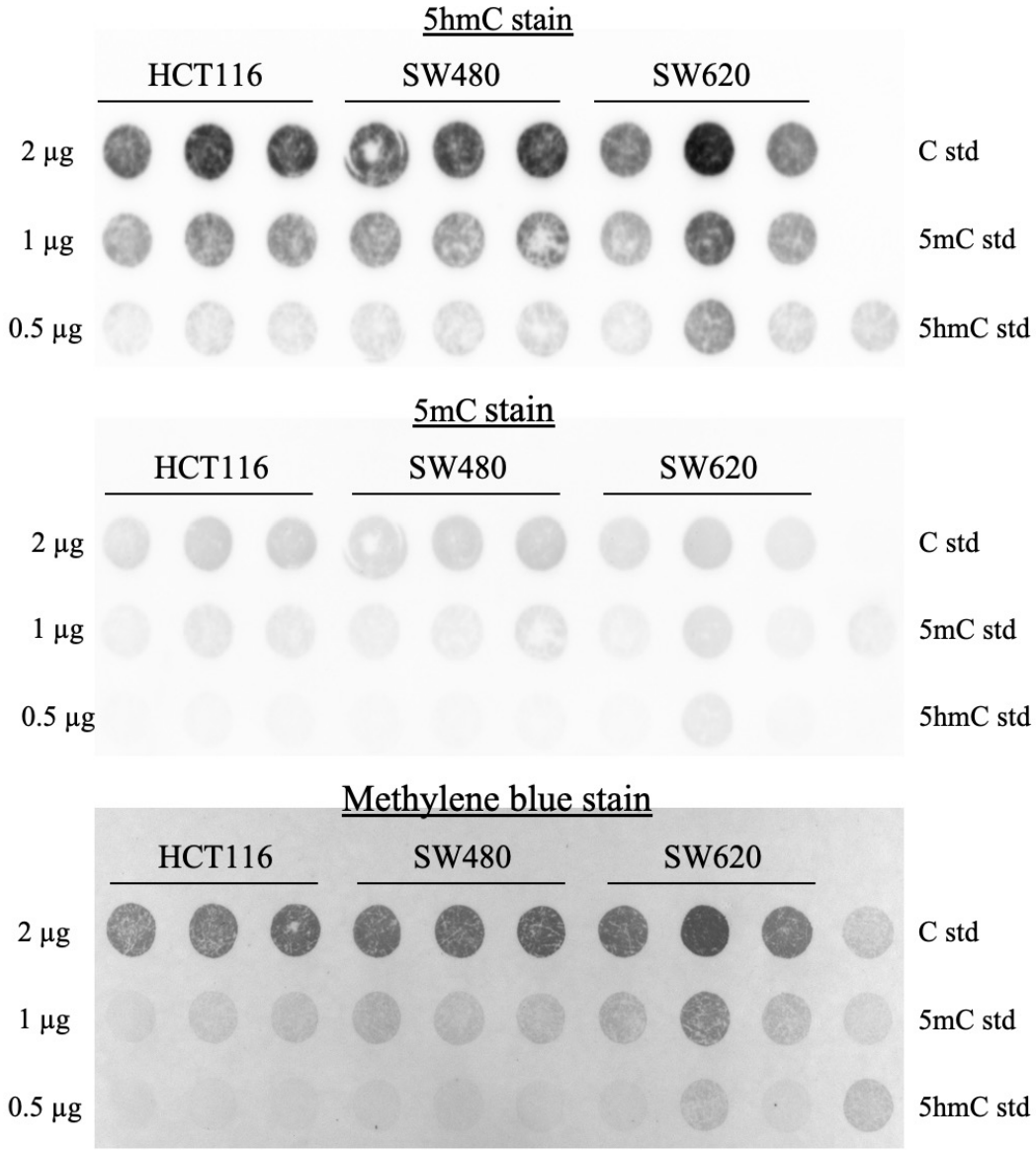


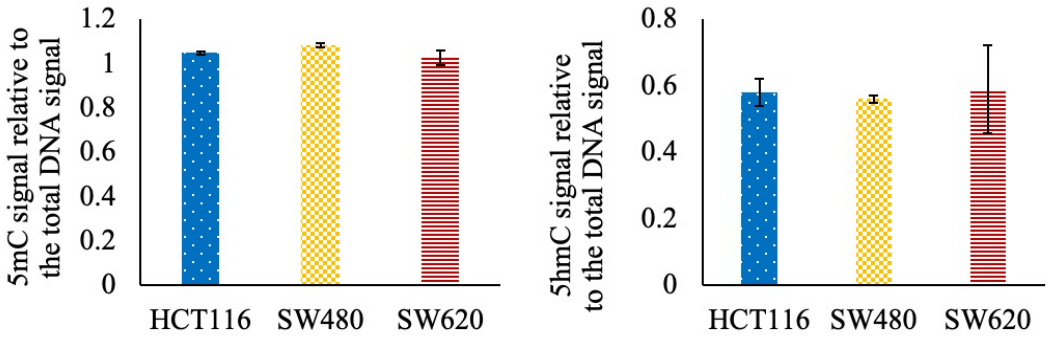
Figure 4.3 The TET2 protein is significantly more abundant in the SW480 cell line compared to the SW620 cells

(A) A representative Western blot image of the assessment of TET1, TET2, TET3 and ß-Tubulin protein levels in SW480 and SW620 cell lines. The bands used for TET1, TET2 and TET3 levles quantification are highlighted by the solid, blue rectangles. The ß-Tubulin band is outlined by the dotted, red rectangles.

(B) The TET protein levels in SW480 and SW620 cell lines. The staining results were normalised to the ß-Tubulin content in each cell line. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$).

A


5hmC stain
HCT116
SW480
SW620
2 µg
1 µg
0.5 µg
C std
5mC std
5hmC std
5mC stain
HCT116
SW480
SW620
2 µg
1 µg
0.5 µg
C std
5mC std
5hmC std
Methylene blue stain
HCT116
SW480
SW620
2 µg
1 µg
0.5 µg
C std
5mC std
5hmC std


5mC signal relative to the total DNA signal
1.2
1
0.8
0.6
0.4
0.2
0
HCT116
SW480
SW620
5hmC signal relative to the total DNA signal
0.8
0.6
0.4
0.2
0
HCT116
SW480
SW620

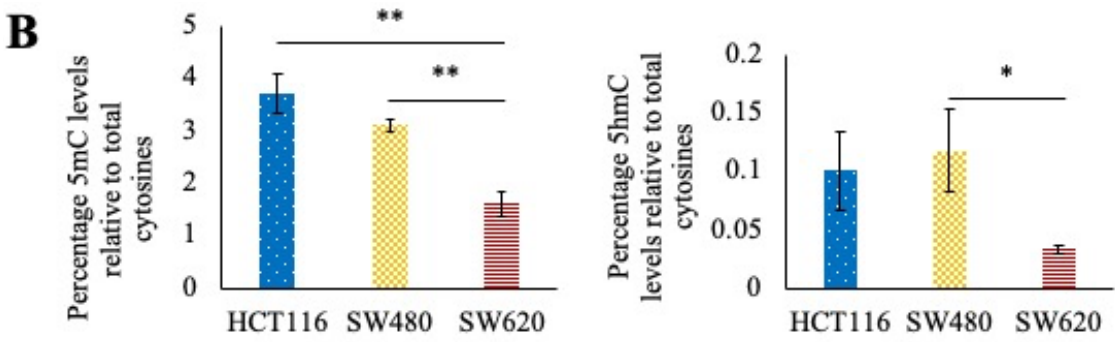


Figure 4.4 The metastatic SW620 cell line has a significantly lower 5mC and 5hmC content than the SW480 primary colorectal cancer cell line
(A) The dot blots showing representative 5hmC, 5mC and methylene blue stains in HCT116, SW480 and SW620 colorectal cancer cell lines and the graphical representation of the results obtained. The error bars represent the standard deviation of three independent replicates.
(B) The mass spectrometry results of 5mC and 5hmC quantification in the three cell lines. The signal obtained for each cytosine modification is displayed as a percentage of signal obtained for all cytosines. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$; ** $p<0.01$).

4.4.2 Establishment of the baseline parameters of proliferation and migration in the SW480 and SW620 cell lines

Following the assessment of TET mRNA and protein levels as well as 5mC and 5hmC content in the two cell lines, their *in vitro* characteristics and properties were investigated. When observed under the light microscope, the two cell lines had similar morphology to the one originally described by Leibovitz and colleagues (Figure 4.5 A) (Leibovitz et al., 1976). The SW480 cells formed large colonies of epithelial-like cells, polygonal in shape with a few protrusions of the cell membrane. The SW620 cells on the other hand formed fewer colonies, with the individual cells either spherical or bipolar with elongated shape. The SW620 cell line was found to have a significantly smaller average diameter (15 µm) than the SW480 cell line (32 µm) (Figure 4.5 B).

Uncontrolled cell proliferation is a hallmark of cancer cells. The rate of proliferation of the SW480 and SW620 cell lines was assessed over 96 hours using trypan blue and haemocytometer-based manual counting (Figure 4.6 A). During the first 24 hours of proliferation, the two cell lines do not show any significant difference in their division

rate. However, over the next 72 hours, SW620 cells were found to proliferate significantly quicker, reaching over double the cell number (1,503,000 cells per mL) at the 96 hours end point compared to the primary line (729,000 cells per mL).

The colony formation assay measures not only the rate of cellular proliferation, but also the ability of single cells in a cell population to form independent colonies. The SW620 cell line was found to have significantly higher clonogenic growth than the SW480 cells (Figure 4.6 B). The number of colonies formed by the SW620 cells was 2.5-fold higher (250 colonies per well) as compared to the SW480 cell line (100 colonies per well). Interestingly, an opposite trend was observed for the colony size. The primary SW480 cells were found to form colonies of a larger size (32 square pixels) than the metastatic SW620 cells (26 square pixels). This difference, albeit small, was found to be significant.

Next, the migration rate of the two cell lines was investigated using the scratch assay. The representative images of the initial (0 hour) and endpoint (48 hours) gap for both cell lines are shown in Figure 4.7 A. The SW480 cell line was found to have a significantly higher migration rate over 48 hours than the SW620 cell line (Figure 4.7 B), closing 36% of the gap after 48 hours compared to 16% for the SW620 cells. This was also true at the 24 hour time point, where the SW480 cells migrated through 23% of the initial gap compared to 7.4% for the SW620 cell line (data not shown).

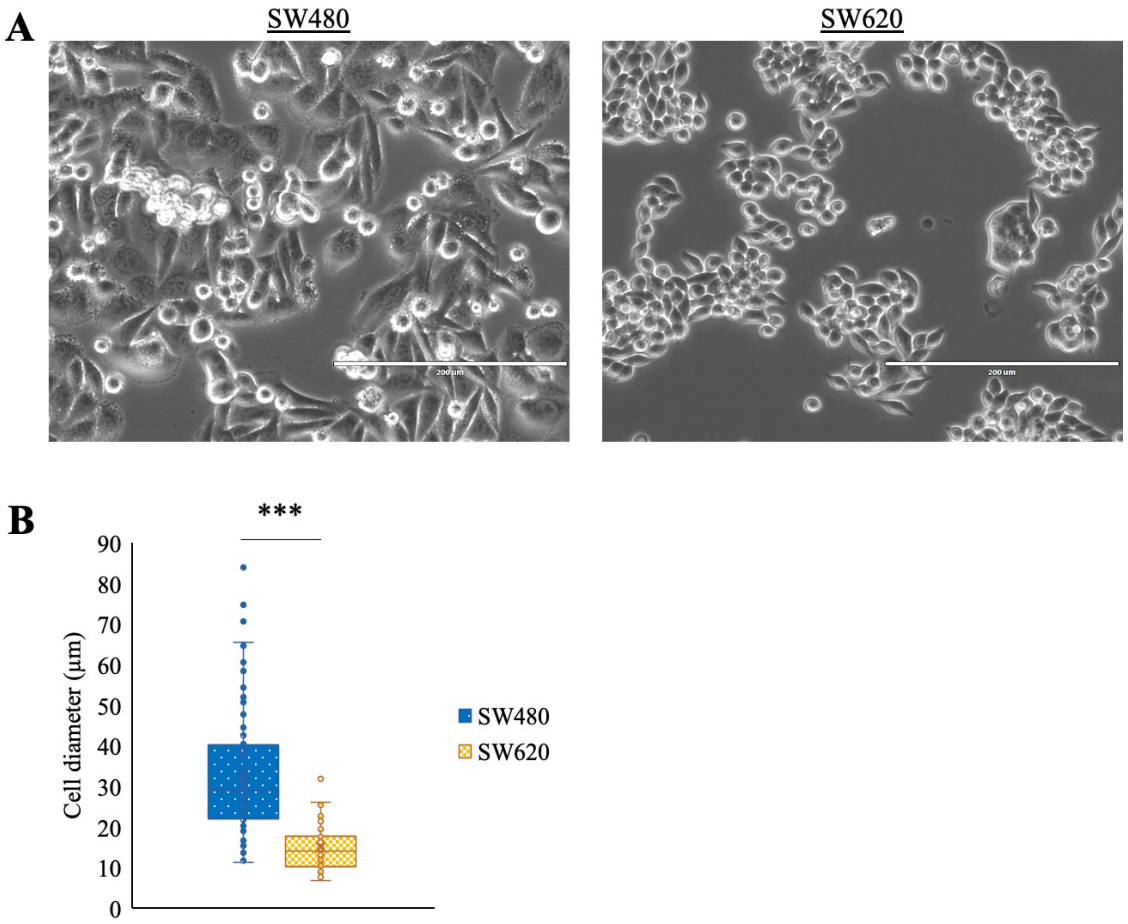


Figure 4.5 The SW620 cell line has a significantly reduced cell size compared to SW480 cell line

(A) The representative images of SW480 and SW620 cell lines taken 48 hours after seeding at 20X magnification, showing the respective cellular morphologies. Scale bar = 200 µm.

(B) The comparison of cell diameter between the SW480 and SW620 cell lines. The error bars represent the standard deviation of one hundred cells measured for each cell line in a single culture dish. The significance was established using an unpaired t-test (*** $p<0.001$).

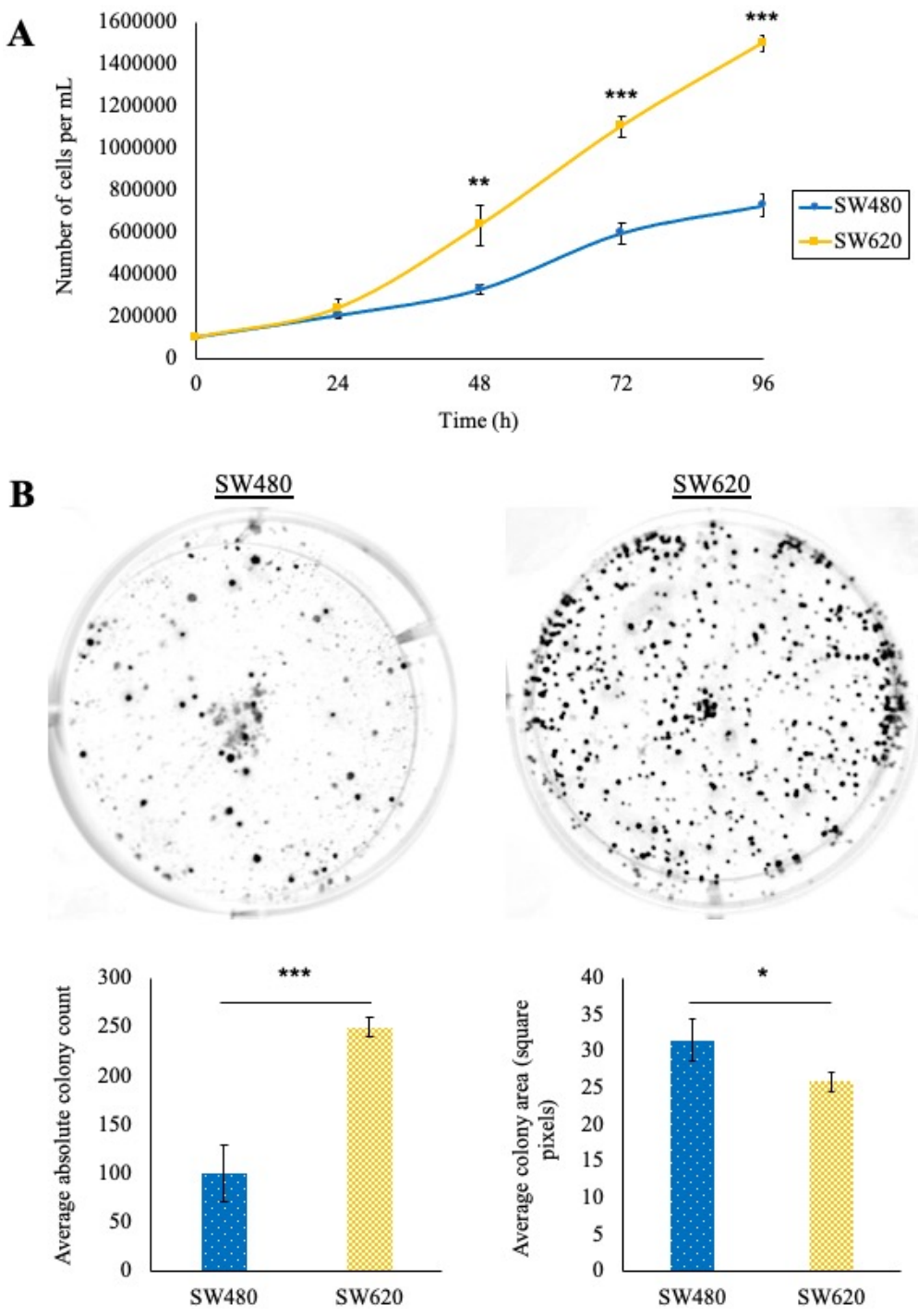


Figure 4.6 The SW620 cell line exhibits a higher proliferation rate and colony formation ability that the SW480 cell line

(A) The cell growth of SW480 and SW620 measured at 24 hour intervals for 96 hours. The error bars represent the standard deviation of three independent replicates. The significance was established using an unpaired t-test (** $p<0.01$; *** $p<0.001$).

(B) The representative images taken of the fixed and stained SW480 and SW620 colonies 12 days after seeding. The bar graphs depict the average colony size and area. The error bars represent the standard deviation of four independent replicates. The significance was established using an unpaired t-test (* $p<0.05$; *** $p<0.001$).

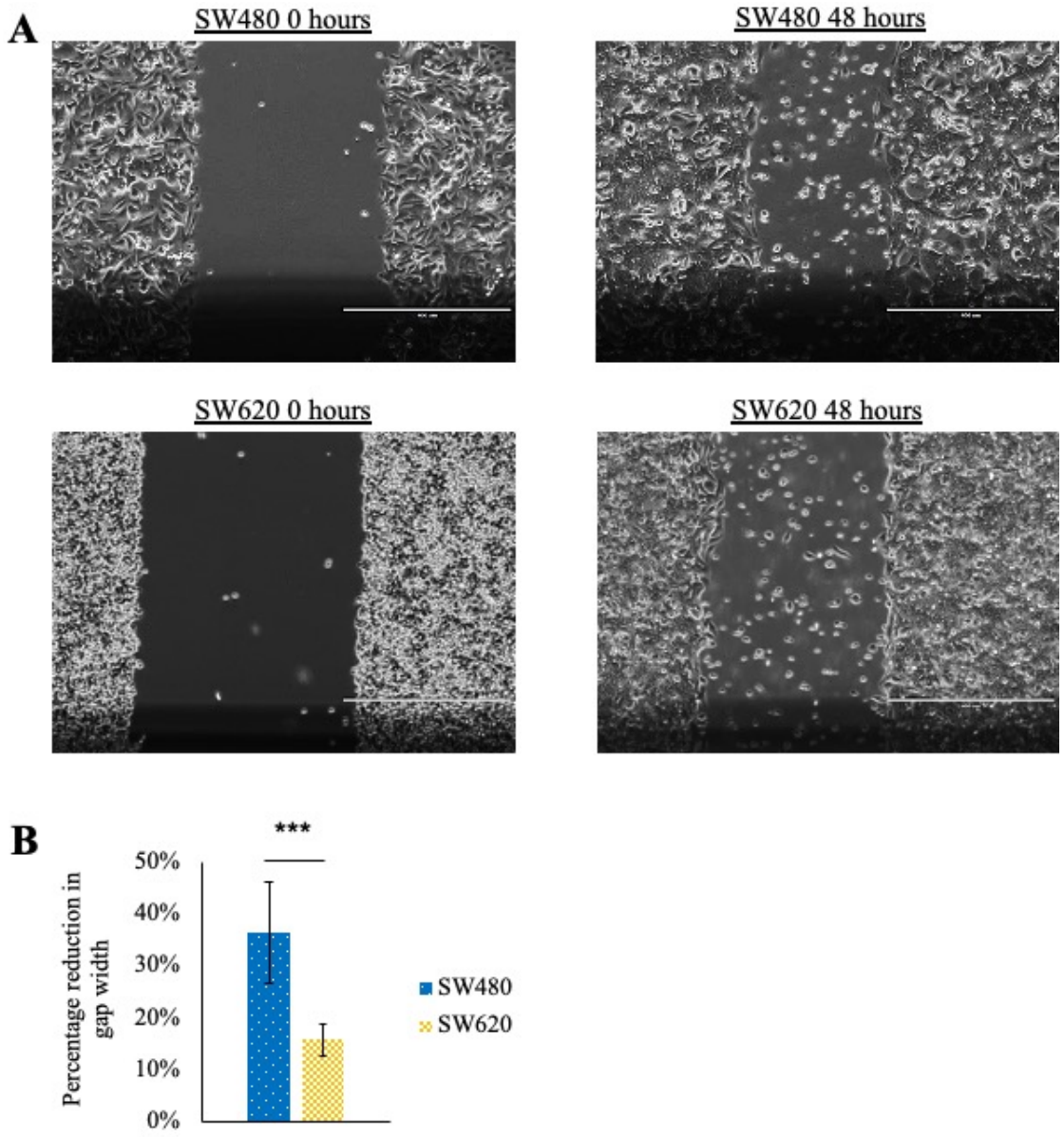


Figure 4.7 The SW480 cell line has a higher migration rate than the SW620 cell line

(A) The representative images of migrating SW480 and SW620 cells taken at 0 hours and 48 hours after the scratch was made, at 10X magnification. Scale bar = 400 µm. (B) The percentage reduction in gap width over 48 hours for SW480 and SW620 cell lines. The error bars represent the standard deviation of eight independent replicates. The significance was established using an unpaired t-test (*** $p<0.001$)

4.4.3 Assessment of the impact of vitamin C, dimethyl 2-oxoglutarate and nickel (II) chloride on TET expression and abundance of 5hmC in SW480 and SW620 cell lines

The analysis of *in vitro* cancer properties of SW480 and SW620 cell lines was followed by the investigation of how the treatment with compounds known to enhance and inhibit the activity of TET enzymes affects the cancer characteristics assessed in the previous section. In order to achieve this, SW480 and SW620 cells were treated with AA (with and without the addition of catalase), DMαKG and $NiCl_2$. The doses used for the initial testing of the compounds for their toxicity (through the MTT assay) were obtained from the literature.

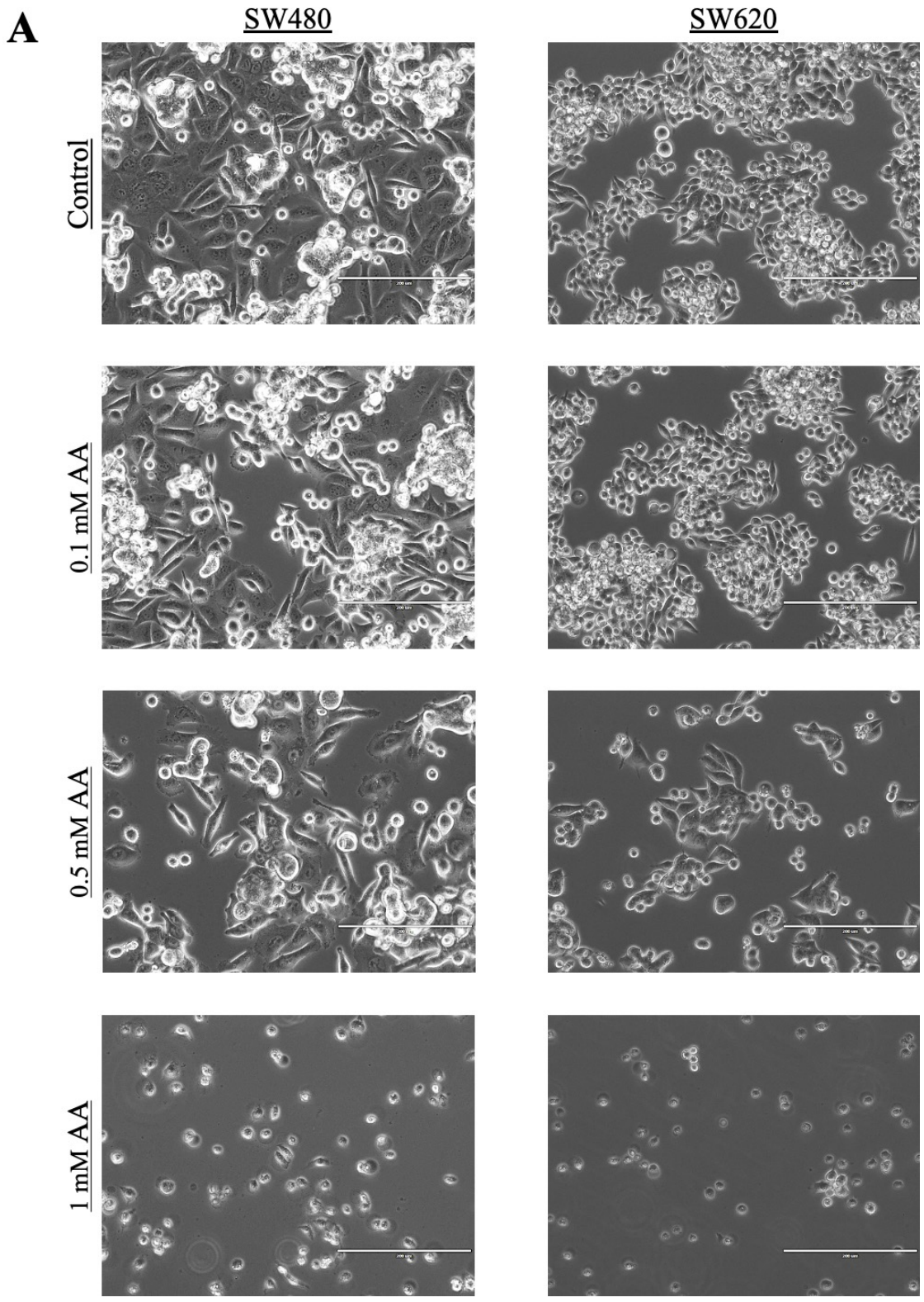
4.4.3.1 The assessment of the effect of ascorbic acid treatment on the *TET* and 5hmC levels in SW480 and SW620 cell lines

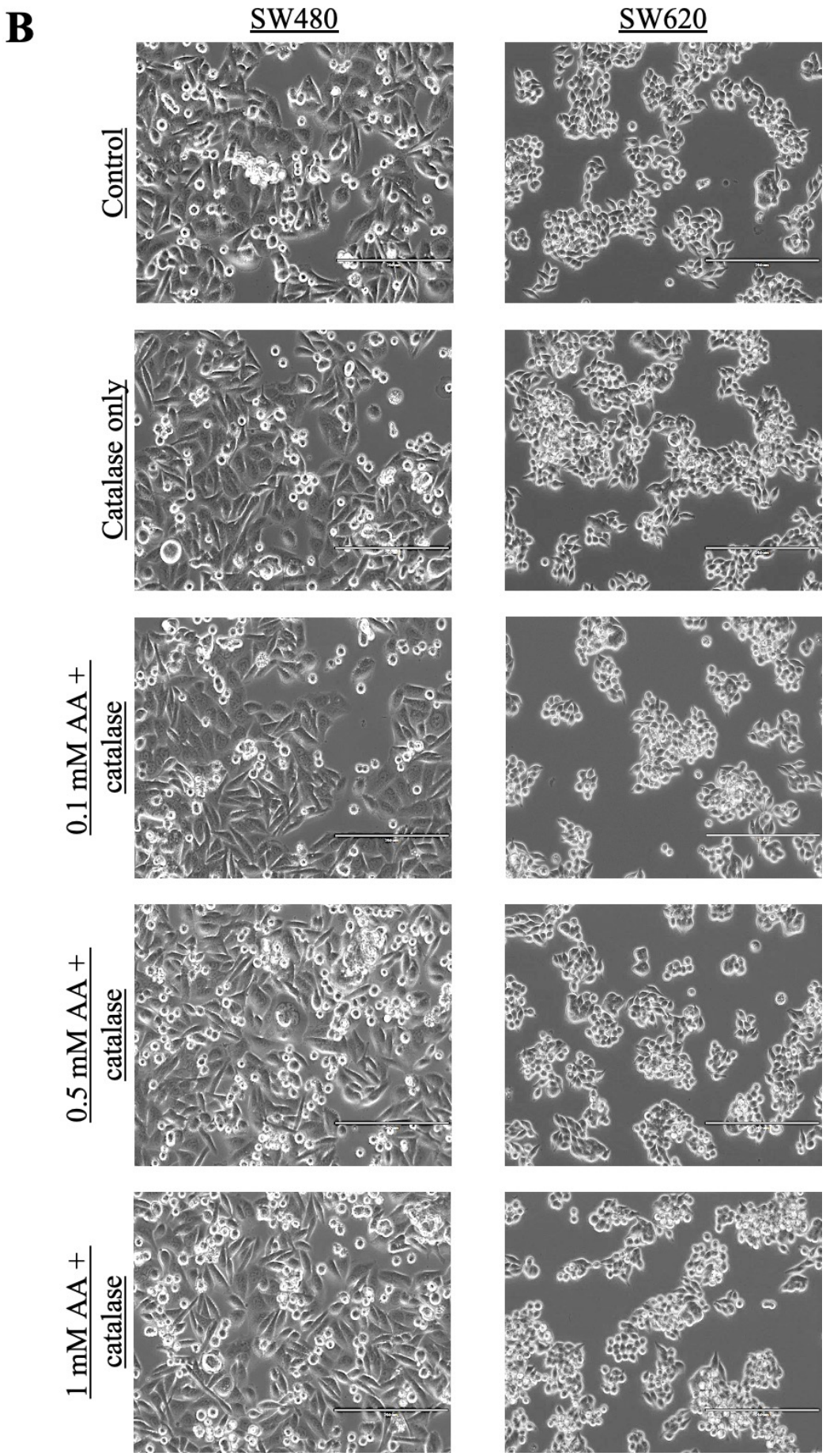
Vitamin C is known to be an essential nutrient that is required for a multitude of physiological processes inside the body. It can, however, also cause cellular toxicity via excessive hydrogen peroxide production (Uetaki et al., 2015). Catalase is known for its ability to scavenge the hydrogen peroxide formed by AA treatment and has been used to differentiate the effects of vitamin C on cancer cells due to epigenetic changes or hydrogen peroxide-mediated cytotoxicity. Based on previous research, a concentration range between 0.1 and 1 mM of AA with and without supplementation of 100 µg/mL catalase solution was used (Blaschke et al., 2013; Shenoy et al., 2017; Shenoy et al., 2019). These treatments were assessed for their cytotoxicity using their morphology and the MTT assay (Figure 4.8). The SW480 cells were found to retain their morphology after 48 hours of treatment with the lowest, 0.1 mM, dose of AA (Figure 4.8 A). The 0.5 mM treatment resulted in more spherical cells being present and a lower cell confluency, while the 1 mM dose resulted in all cells losing their wild type morphology and a drastically reduced cell number.

The same pattern was observed for the SW620 cells, except the effect of most doses appeared to be more pronounced. These morphological changes were prevented from occurring in both cell lines by the addition of the catalase enzyme (Figure 4.8 B). This was true even for the highest, 1 mM AA concentration. The morphological changes

observed were mirrored by the MTT assay results. The 0.5 mM and 1 mM concentrations of AA led to large reductions in cell viability of the SW480 (82 and 97% reduction, respectively) and SW620 (80 and 99% reductions, respectively) cell lines while the 0.1 mM treatment had no apparent effect in that regard (Figure 4.8 C). The reduction in cell viability was rescued by the addition of catalase (Figure 4.8 D). Due to the fact that catalase prevented cytotoxicity at 1 mM AA concertation, the 0.5 mM of vitamin C with catalase was not included in the further assays.

While it is known that vitamin C affects the activity of TET enzymes, it is unclear whether it affects the levels of their transcripts. The levels of *TET* mRNAs were assessed following a 48 hour AA treatment with and without the presence of catalase. In the SW480 cell line, *TET1* and *TET2* levels were elevated, albeit not significantly, in the AA treated SW480 cells (Figure 4.9 A). There were no such changes seen in the samples supplemented with catalase. *TET3* levels on the other hand appeared to be reduced in all treated samples, although none of these changes were found to be significant. In the SW620 cell line, none of the *TET* transcripts showed significant changes in their levels following the treatments (Figure 4.9 B).


A
SW480
SW620
Control
0.1 mM AA
0.5 mM AA
1 mM AA


B
SW480
SW620
Control
Catalase only
0.1 mM AA + catalase
0.5 mM AA + catalase
1 mM AA + catalase

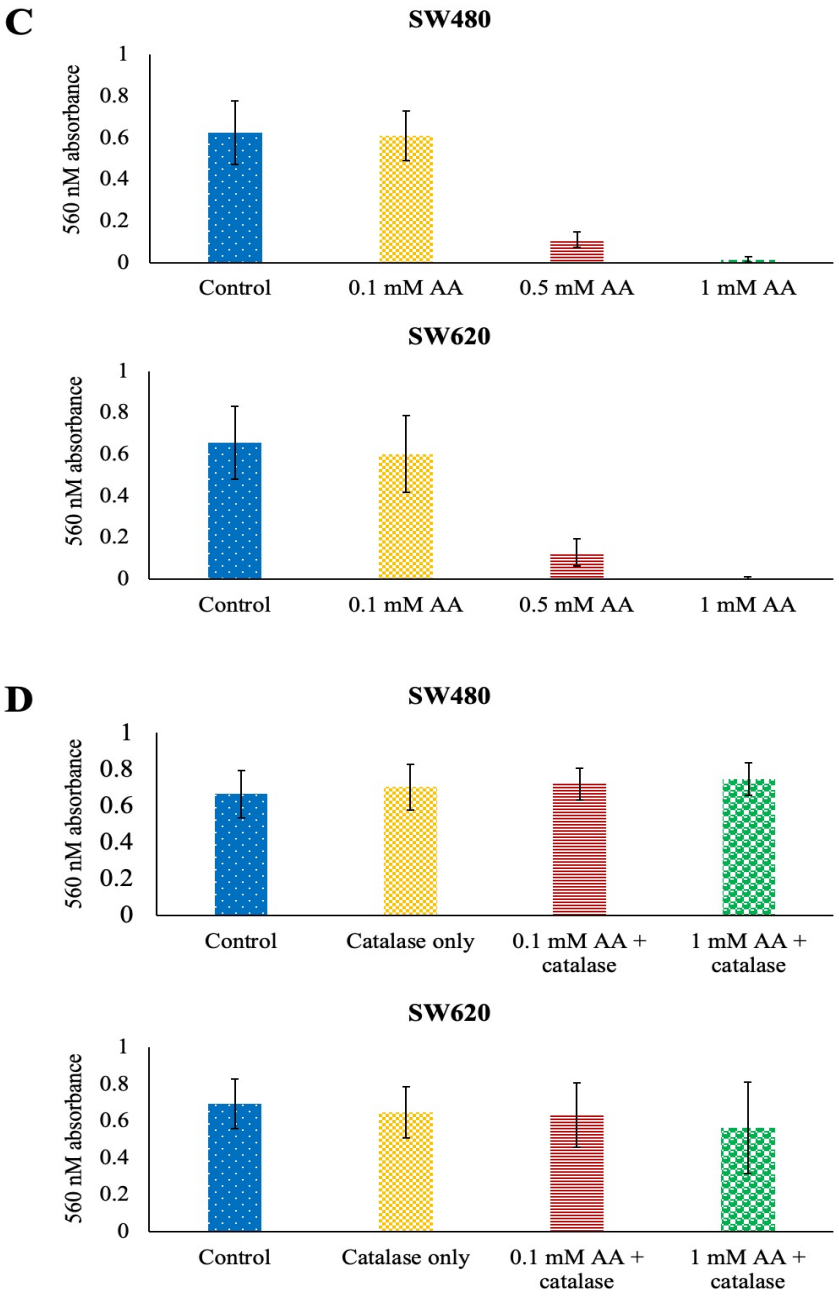


Figure 4.8 The cytotoxicity of ascorbic acid treatment is rescued by addition of catalase

(A) The morphology of the two cell lines after a 48 hour treatment with the indicated doses of ascorbic acid (AA). The images were taken at 20X magnification. Scale bar = 200 µm.

(B) The morphology of the two cell lines after a 48 hour treatment with the indicated doses of AA supplemented with the catalase enzyme. The images were taken at 20X magnification. Scale bar = 200 µm.

(C) and (D) The results of the MTT assay following a 48 treatment of the two cell lines with the indicated doses of AA alone or AA supplemented with catalase, respectively. The 560 nM absorbance indicates the degree of cell viability. The error bars represent the standard deviation of three independent replicates.

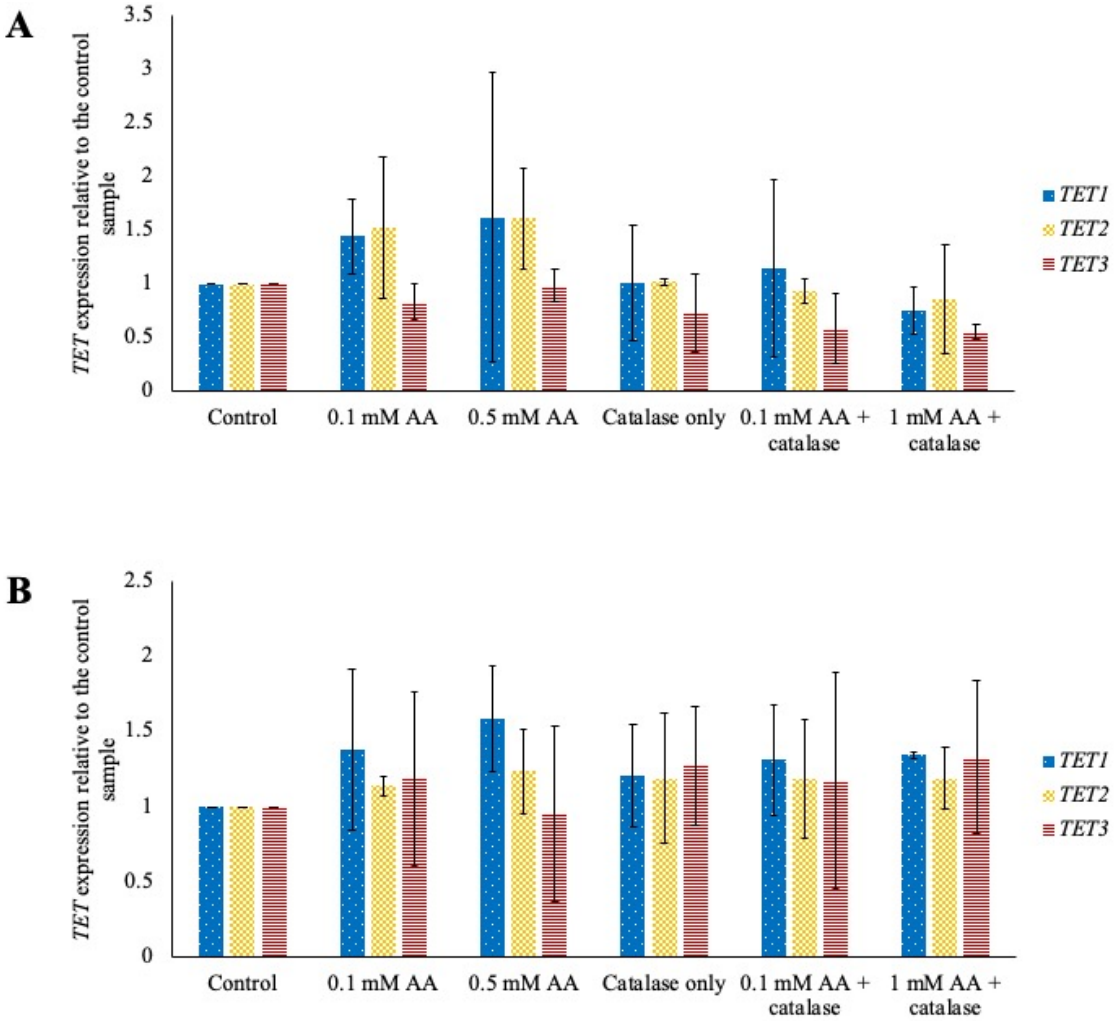


Figure 4.9 The levels of *TET* mRNAs remain largely unaffected following a 48 hour ascorbic acid treatment of SW480 and SW620 cells with and without catalase supplementation

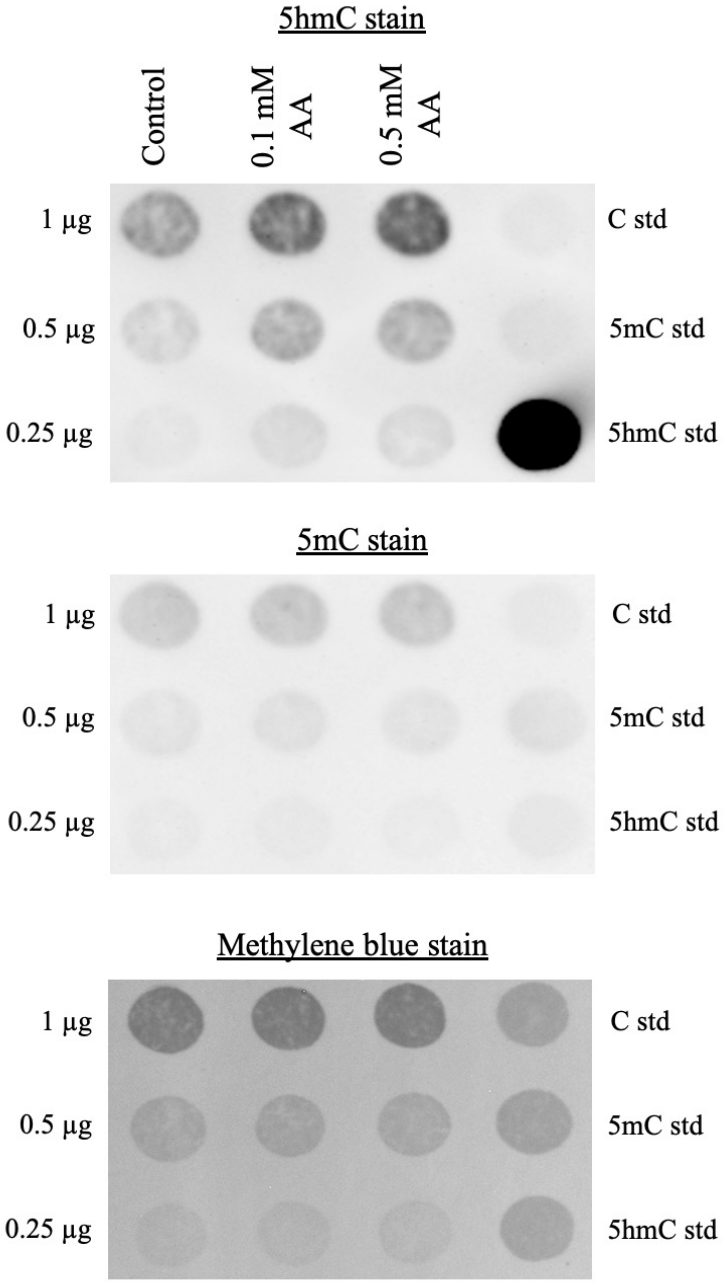
The expression of the *TET* transcripts in (A) SW480 and (B) SW620 cell lines treated with ascorbic acid (AA) and catalase. The *TET* mRNA levels were normalised to *ACTB* reference gene. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison.

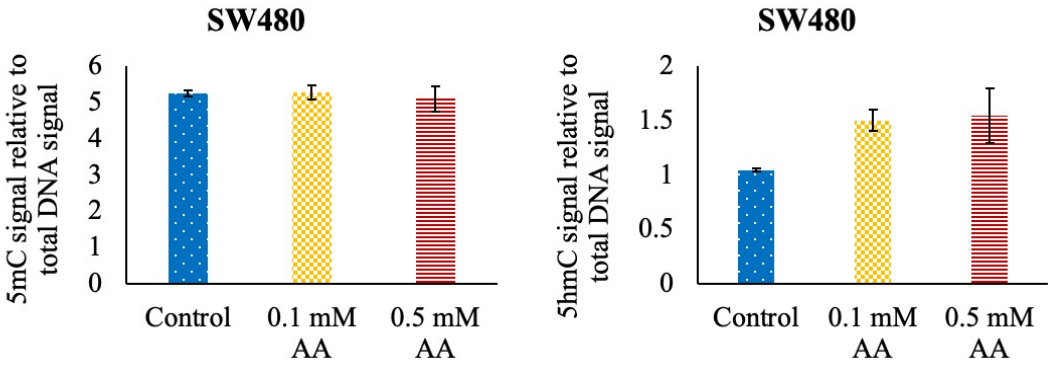
The ability of AA to increase the global 5hmC levels has been demonstrated in numerous cell line types. The abundance of both 5mC and 5hmC in the SW480 and SW620 cell lines following incubation with either vitamin C on its own or in conjunction with catalase was assessed using both the dot blot staining (Figure 4.10) and mass spectrometry (Figure 4.11). The 48 hour incubation of both cell lines with 0.1 and 0.5 mM of vitamin C resulted in a mild upregulation of global 5hmC levels in both cell lines with a greater upregulation in the SW480 cells (1.50- and 1.54-fold,

respectively) (Figure 4.10 A) than SW620 cells (1.03 and 1.06, respectively) (Figure 4.10 B) compared to the untreated samples. The global levels of 5mC were unchanged in both cell lines. Similarly, AA treatment of the cells in the presence of catalase resulted in an increase in 5hmC content in both SW480 (Figure 4.10 C) and SW620 (Figure 4.10 D). The 0.1, 0.5 and 1 mM AA supplemented with catalase treatments resulted in 5.1-, 4.2- and 3.6-fold upregulation of 5hmC content in SW480 cells compared to untreated control and 4.2-, 3.5- and 3.6-fold increase in SW620 cells. Intriguingly, the increase in 5hmC levels compared to untreated and catalase only controls was much larger than was observed for the AA on its own, despite catalase treatment itself not having an effect on global 5hmC content. The abundance of 5mC was again found to be unchanged by the AA and catalase treatments.

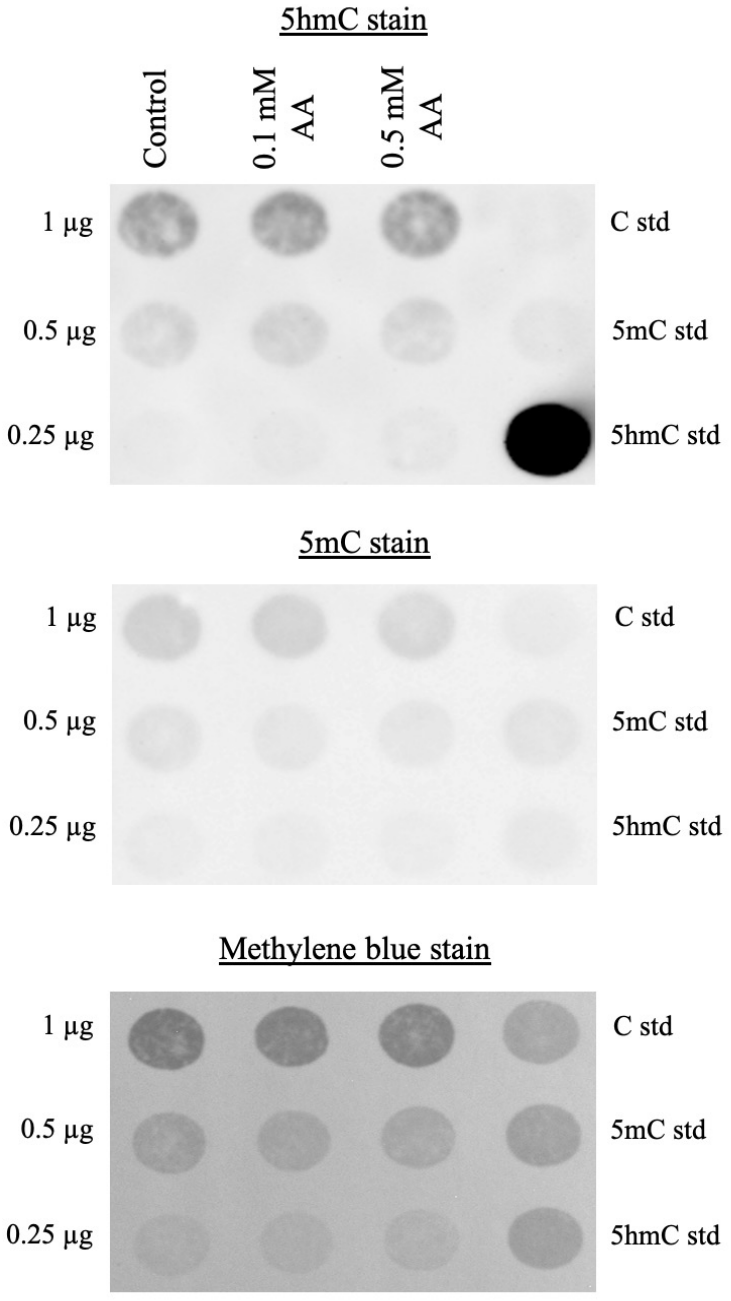
Mass spectrometry measurements of cytosine modifications following AA treatment contradicted the results outlined above (Figure 4.11 A). The 5hmC levels were mildly upregulated in both cell lines following the incubation with the two selected doses, although this did not reach significance in either of the cell lines. In line with the dot blot results, the 5mC levels were unchanged in both SW480 and SW620 cells. When the vitamin C was combined with catalase, an increase in 5hmC levels was observed with the 1 mm vitamin C dose in both cell lines, however this was determined to not be significant. On the other hand, little change in the abundance of 5hmC was seen with the 0.1 mM dose in either cell line. Just as in the earlier observations, the 5mC levels showed little change with a small increase in the samples treated with 1 mM AA and catalase. The mass spectrometry results therefore indicate that AA treatment does not induce the expected increase in global 5hmC levles in neither the SW480 nor SW620 cells.

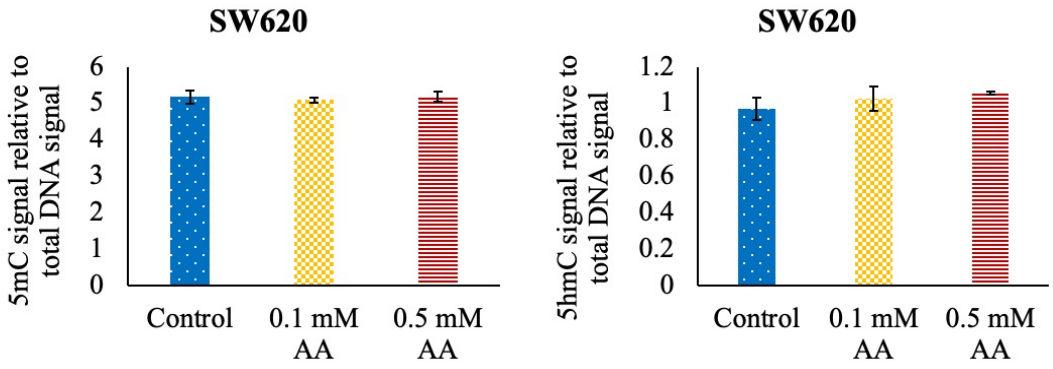
A


5hmC stain
Control
0.1 mM AA
0.5 mM AA
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
5mC stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
Methylene blue stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std

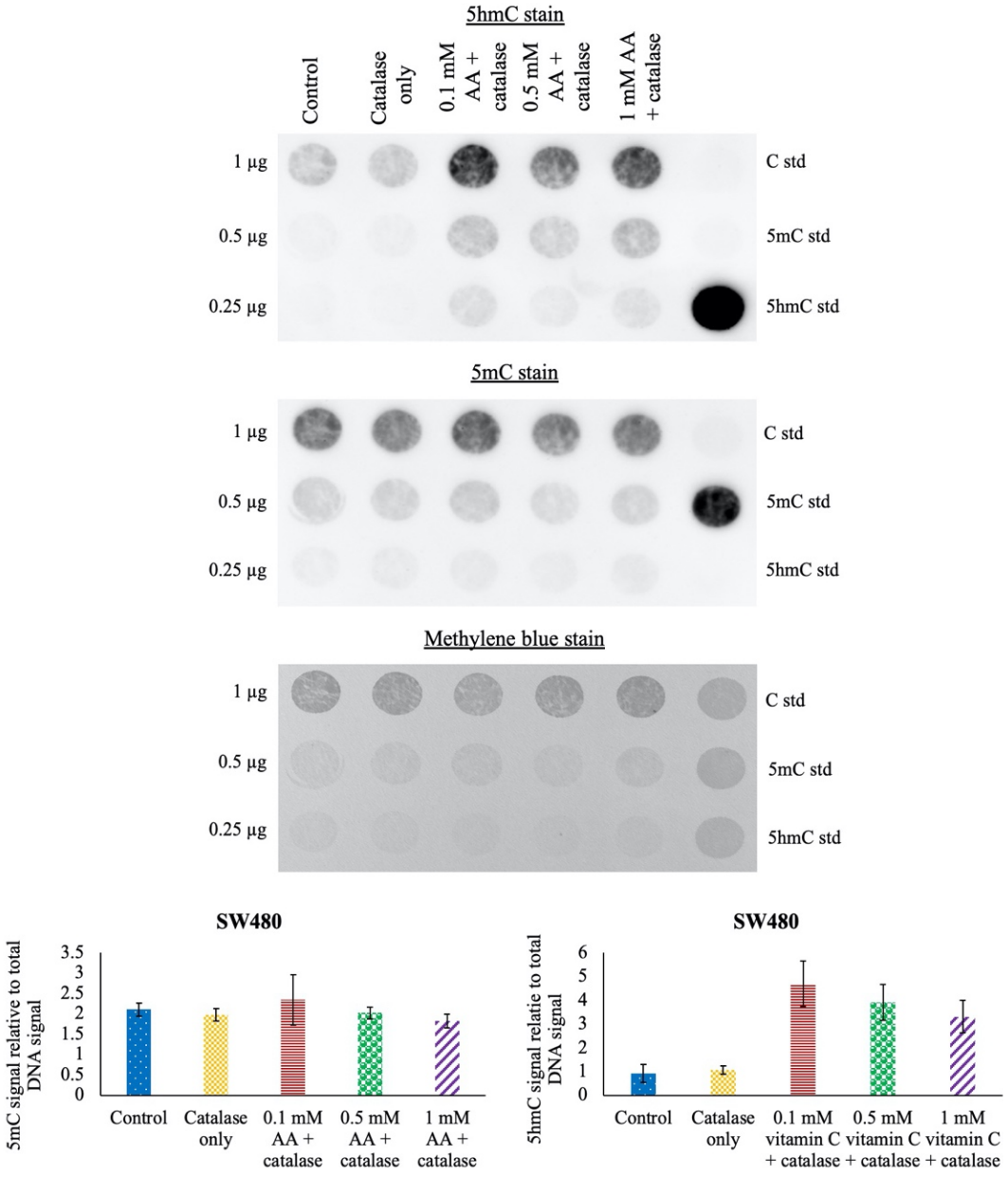

SW480
5mC signal relative to total DNA signal
6
5
4
3
2
1
0
Control
0.1 mM AA
0.5 mM AA
SW480
5hmC signal relative to total DNA signal
2
1.5
1
0.5
0
Control
0.1 mM AA
0.5 mM AA

B


5hmC stain
Control
0.1 mM AA
0.5 mM AA
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
5mC stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
Methylene blue stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std


SW620
5mC signal relative to total DNA signal
6
5
4
3
2
1
0
Control
0.1 mM AA
0.5 mM AA
SW620
5hmC signal relative to total DNA signal
1.2
1
0.8
0.6
0.4
0.2
0
Control
0.1 mM AA
0.5 mM AA

C


5hmC stain
Control
Catalase only
0.1 mM AA + catalase
0.5 mM AA + catalase
1 mM AA + catalase
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
5mC stain
Methylene blue stain
SW480
5mC signal relative to total DNA signal
Control
Catalase only
0.1 mM AA + catalase
0.5 mM AA + catalase
1 mM AA + catalase
SW480
5hmC signal relatie to total DNA signal
0.1 mM vitamin C + catalase
0.5 mM vitamin C + catalase
1 mM vitamin C + catalase

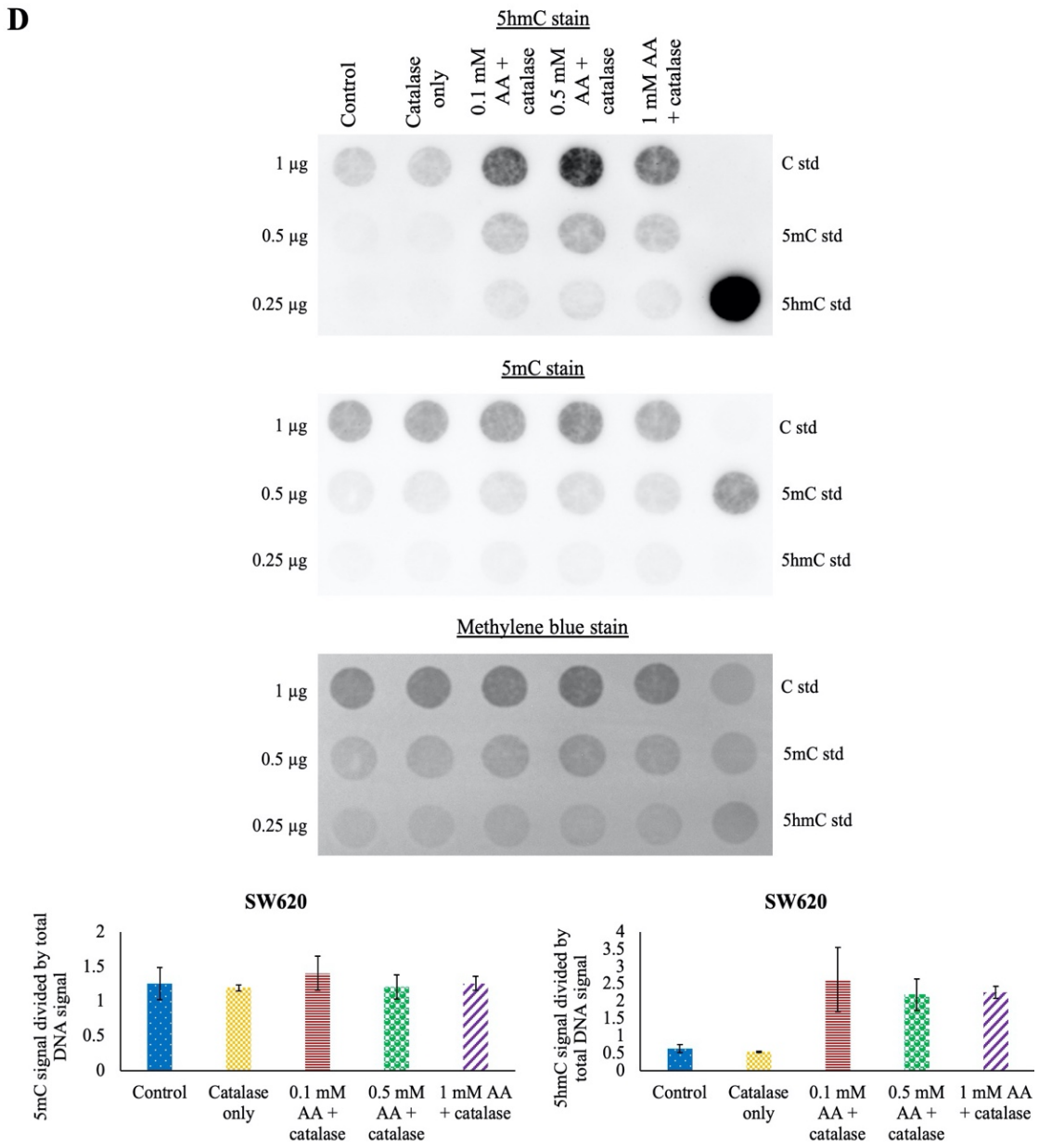


Figure 4.10 Ascorbic acid with and without the presence of catalase causes a small upregulation of global 5hmC, but not 5mC, levels in SW480 and SW620 cell lines following a 48 hour treatment

(A) and (B) The representative staining results of dot blot assessment of 5hmC and 5mC levels in SW480 and SW620 cell lines, respectively, following a 48 hour incubation with indicated doses of ascorbic acid (AA). The errors bars represent the standard deviation of three independent replicates.

(C) and (D) The representative staining results of dot blot assessment of 5hmC and 5mC levels in SW480 and SW620 cell lines, respectively, following a 48 hour incubation with catalase only and indicated doses of AA with catalase. The errors bars represent the standard deviation of three independent replicates.

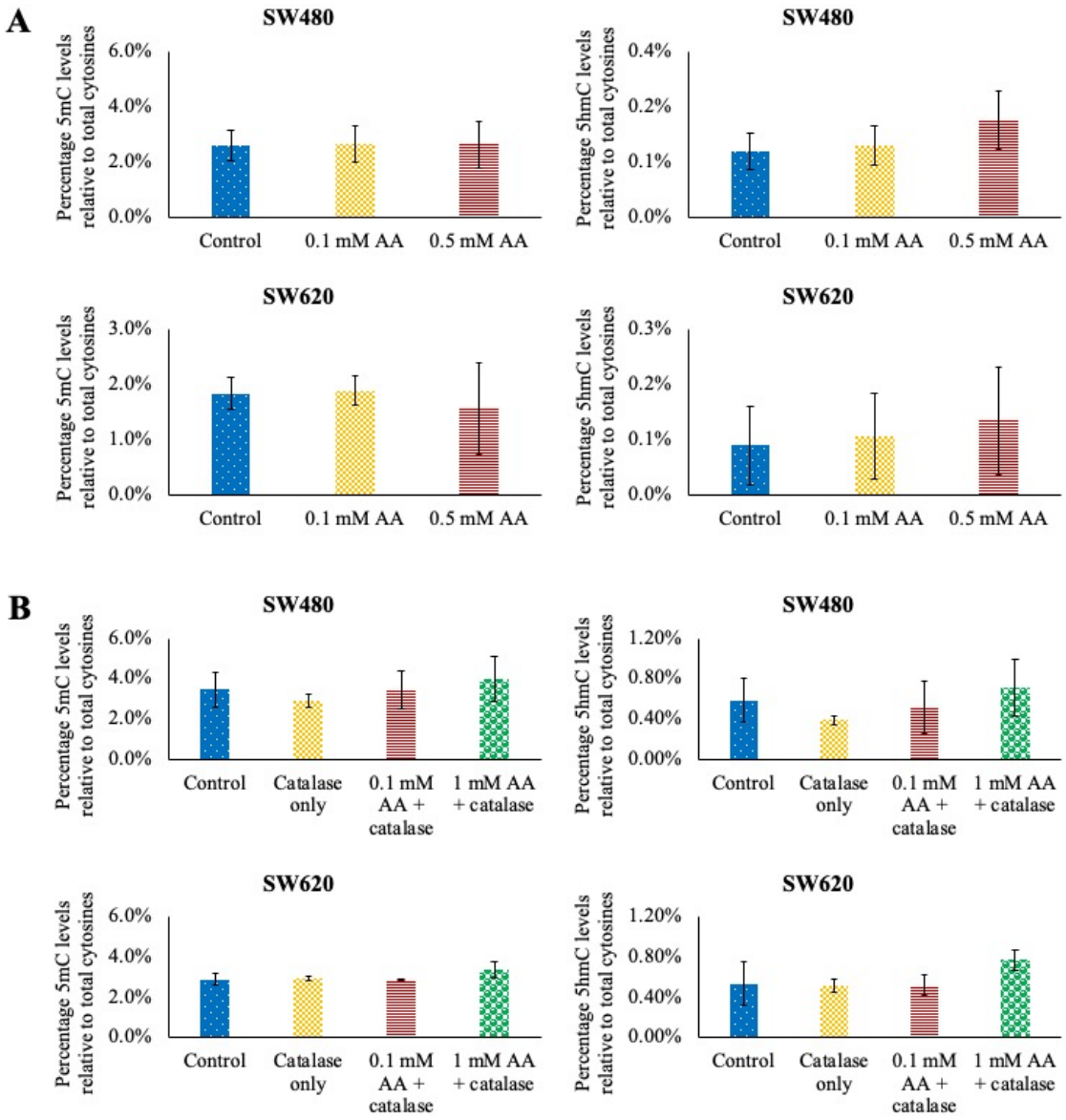


Figure 4.11 Mass spectrometry assessment of the global 5mC and 5hmC levels in SW480 and SW620 cell lines following a 48 hour ascorbic acid treatment
(A) and (B) The results of mass spectrometry assessment of 5hmC and 5mC levels in SW480 and SW620 cell lines following a 48 hour incubation with indicated doses of ascorbic acid (AA) or AA supplemented with catalase, respectively. The errors bars represent the standard deviation of six (for AA) or three (for AA with catalase) independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison.

4.4.3.2 The assessment of the effect of nickel (II) chloride and dimethyl 2-oxoglutarate treatment on the *TET* and 5hmC levels in SW480 and SW620 cell lines

The experiments aiming at elucidating the possible role of AA-induced increase in 5hmC levels on the *in vitro* cancer characteristics of SW480 and SW620 cell lines were followed by similar analysis with $NiCl_2$ and DMαKG. This would allow for direct comparison of results obtained with vitamin C to those with another compound known to increase 5hmC levels, but through a different mechanism. Should the results obtained be similar, this will make it more likely that they are due to the increase of 5hmC levels rather than any off-target effects. On the other hand, one would then expect to see the opposite trend following the treatment of cells with $NiCl_2$ given its inhibitory action on the TET enzymes.

The doses of the two compounds previously used in the literature were tested for their cytotoxicity in the SW480 and SW620 cells through the MTT assay (Figure 4.12). The three concentrations of DMαKG tested did not result in increased cell death in either of the cell lines following 48 hours of treatment (Figure 4.12 A). On the other hand, the highest dose (200 μM) of $NiCl_2$ lowered the cell viability by 46% and 50% in SW480 and SW620 cell line, respectively, compared to untreated control (Figure 4.12 B). Given the high acidity of DMαKG, it was speculated that one might be able to use higher doses of it in culture if it is supplemented with a suitable buffer. In order to test that, two higher concentrations of DMαKG (10 mM and 25 mM) with and without 10 mM HEPES supplementation were assessed for their cytotoxic properties. It was found that the 10 mM DMαKG dose resulted in a small degree of cell death in both the SW480 (14%) (Figure 4.12 C) and the SW620 (22%) (Figure 4.12 D) cell lines. When the concentration of the compound was increased to 25 mM, the cell viability was reduced by 69% in the primary cell line and 93% in the metastatic counterpart. The addition of HEPES to both doses did not results in a rescue of viability in either cell line.

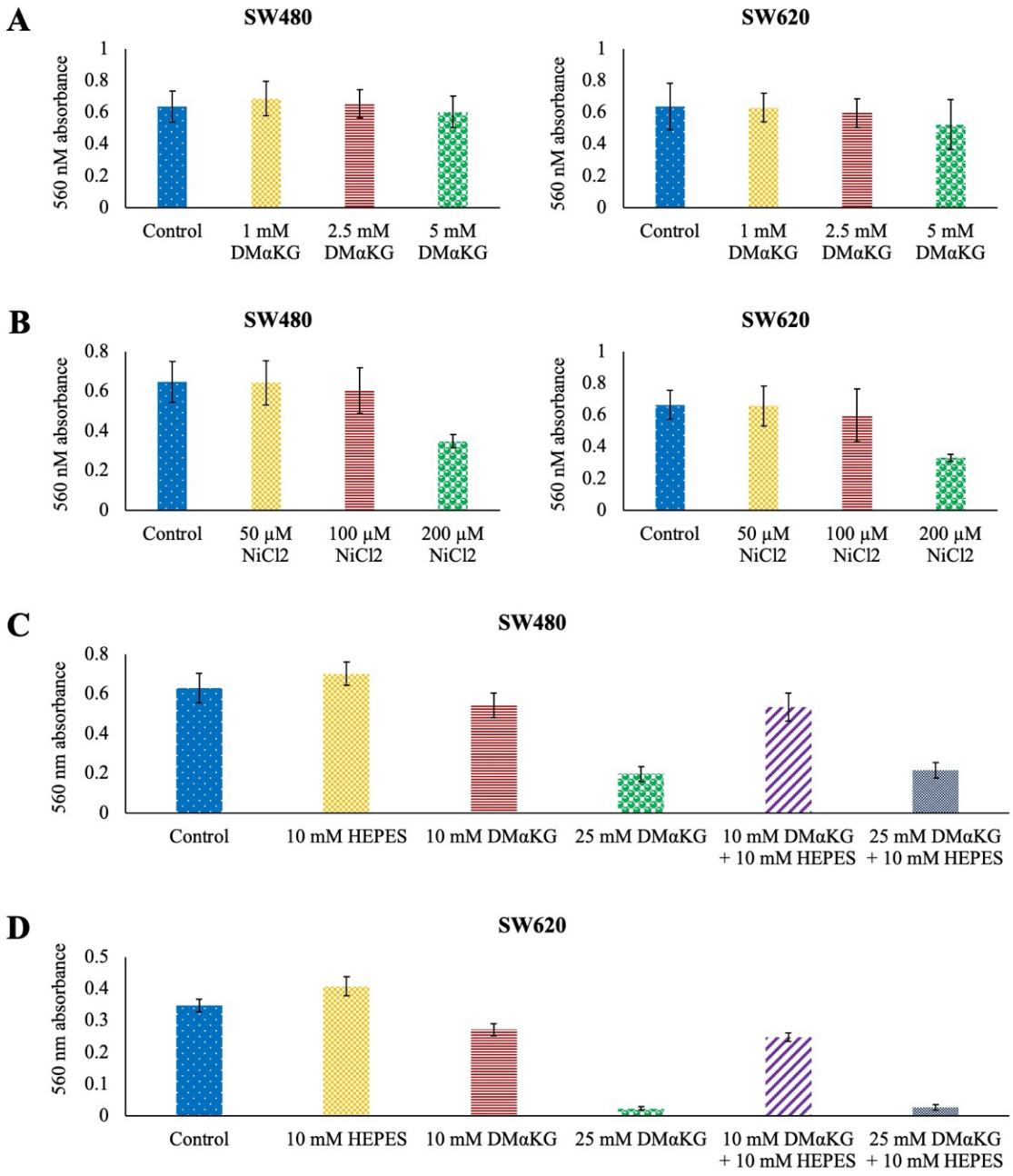


Figure 4.12 High doses of dimethyl alpha-ketoglutarate and nickel (II) chloride are cytotoxic in SW480 and SW620 cell lines

(A) and (B) The results of the MTT assay following a 48 hour treatment of the SW480 and SW620 cell lines with dimethyl alpha-ketoglutarate (DMαKG) or nickel (II) chloride ($NiCl_2$), respectively. The 560 nM absorbance indicates the degree of cell viability. The error bars represent the standard deviation of three independent replicates. (C) and (D) The results of the MTT assay following a 48 hour treatment of the SW480 and SW620 cell lines, respectively, with higher doses of DMαKG with and without HEPES supplementation. The 560 nM absorbance indicates the degree of cell viability. The error bars represent the standard deviation of three independent replicates.

The results of the MTT assay described above led to the selection of following doses of the two compounds for subsequent experiments: 1 mM and 5 mM of DMαKG as well as 50 µM and 100 µM of $NiCl_2$. These concentrations were used to assess the effect of these treatments on the levels of *TET* transcripts (Figure 4.13). The DMαKG resulted in mild upregulation of *TET1* and *TET2* transcripts in both cell lines (Figure 4.13 A). The expression of *TET3* remained largely unchanged with its levels mildly reduced in treated SW480 cells. None of the above changes were found to be significant. The treatment of SW480 cells with $NiCl_2$ did not result in significant changes in *TET* expression overall (Figure 4.13 B). The only significant changes observed in SW620 cells following the treatments were the 2.2-fold and 1.9-fold upregulations of the *TET1* transcript in cells incubated with 100 µM $NiCl_2$ relative to the control and 50 µM $NiCl_2$ samples, respectively.

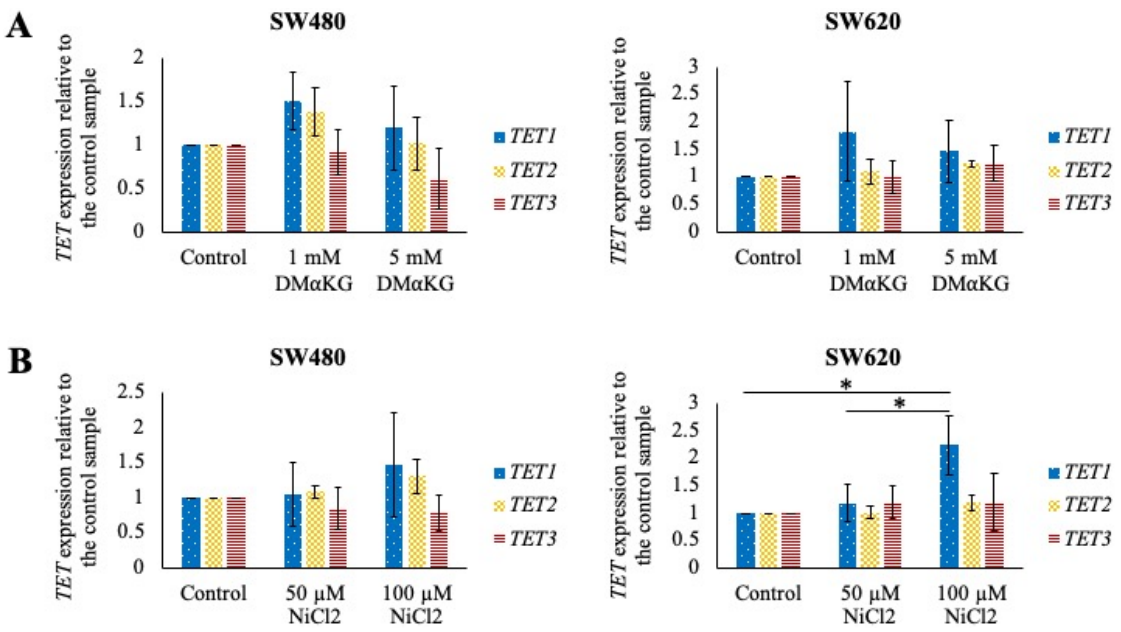


Figure 4.13 The levels of *TET* mRNAs remain largely unaffected following a 48 hour dimethyl alpha-ketoglutarate and nickel chloride treatments of SW480 and SW620 cells

The expression of the *TET* transcripts in (A) dimethyl alpha-ketoglutarate (DMαKG) and (B) nickel (II) chloride ($NiCl_2$) treated SW480 and SW620 cells. The *TET* mRNA levels were normalised to *ACTB* reference gene. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$).

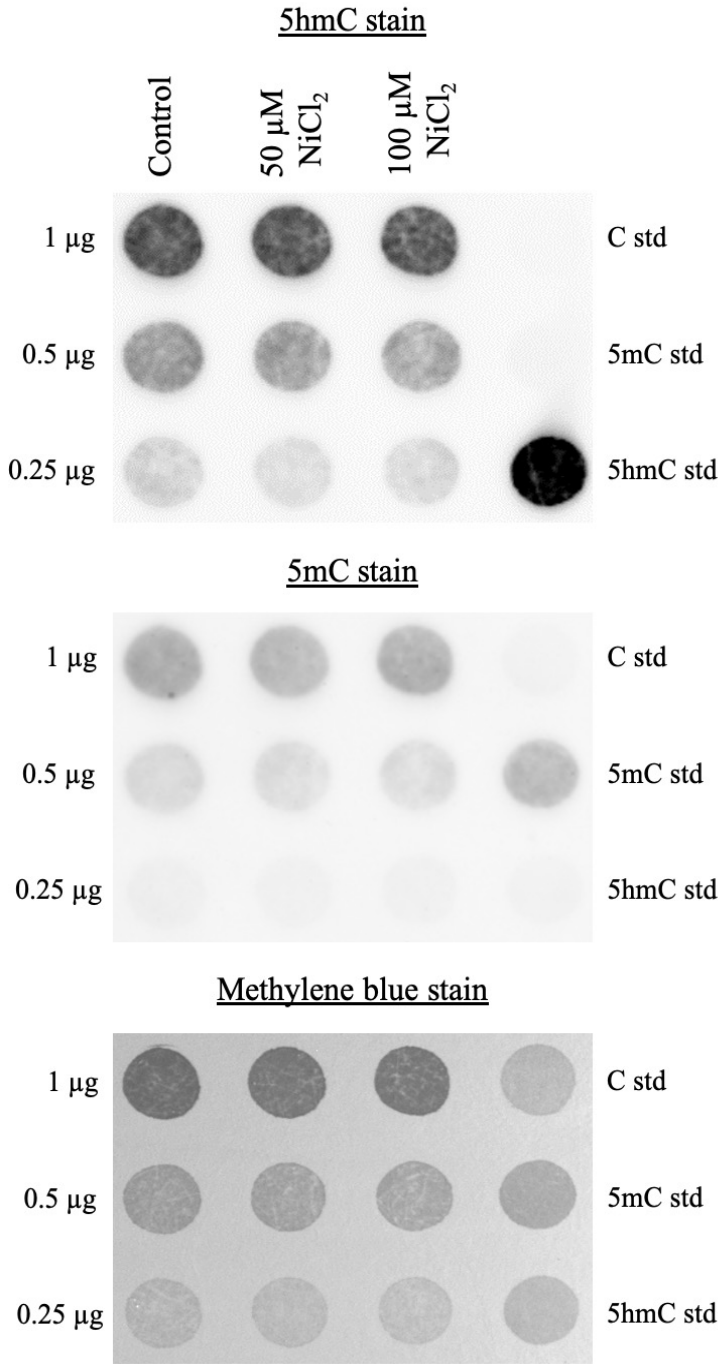
The 5hmC and 5mC content of the $NiCl_2$- and DMαKG-treated SW480 and SW620 cells was next assessed using the dot blot technique (Figure 4.14). The SW480 cell line DNA showed no changes in neither 5hmC nor 5mC content following the treatment with either of the $NiCl_2$ doses (Figure 4.14 A). The same pattern was observed with the SW620 cell line DNA, albeit a mild reduction of both 5mC and 5hmC levels can be seen in the 50 μM $NiCl_2$-treated sample (Figure 4.14 B). The SW480 cells treated with the selected doses of DMαKG showed no changes in global 5mC levels while the abundance of 5hmC was mildly lowered by the 1mM (9%) and 5mM (14%) treatments compared to the untreated control (Figure 4.14 C). Following the same treatment, the SW620 cells' DNA showed no changes in global 5mC and 5hmC levels (Figure 4.14 D).

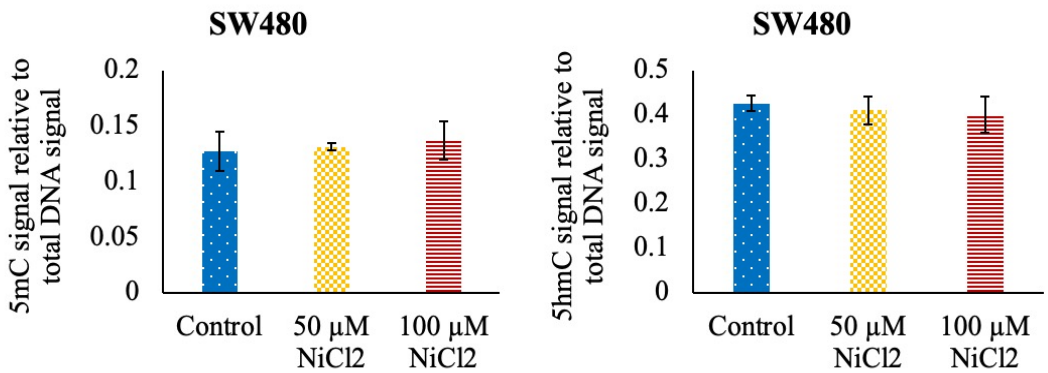
These results were subsequently confirmed using the in-house mass spectrometry technique (Figure 4.15). The $NiCl_2$ treatment resulted in no changes in 5mC and 5hmC levels in either cell line (Figure 4.15 A). This is in stark contrast to the expected result given that $NiCl_2$ is known to inhibit the TET enzymes' function. There were no significant changes in 5mC and 5hmC levels following the DMαKG treatment in neither of the two cell lines (Figure 4.15 B).

In order to test whether a higher dose of DMαKG might be required to produce an increase in 5hmC levels in the two CRC cell lines, the cells were incubated with 10 mM dimethyl alpha ketoglutarate with and without the presence of 10 mM HEPES solution. HEPES was used in an attempt to buffer the highly acidic DMαKG. Following a 48 hour incubation period, the DNA from both cell lines was assessed for 5mC and 5hmC content using the dot blot technique (Figure 4.16). It was found that the increased dose of DMαKG did not result in the expected increase in 5hmC levels in neither the SW480 (Figure 4.16 A) nor SW620 (Figure 4.16 B) cell line. The abundance of 5mC also remained unchanged.

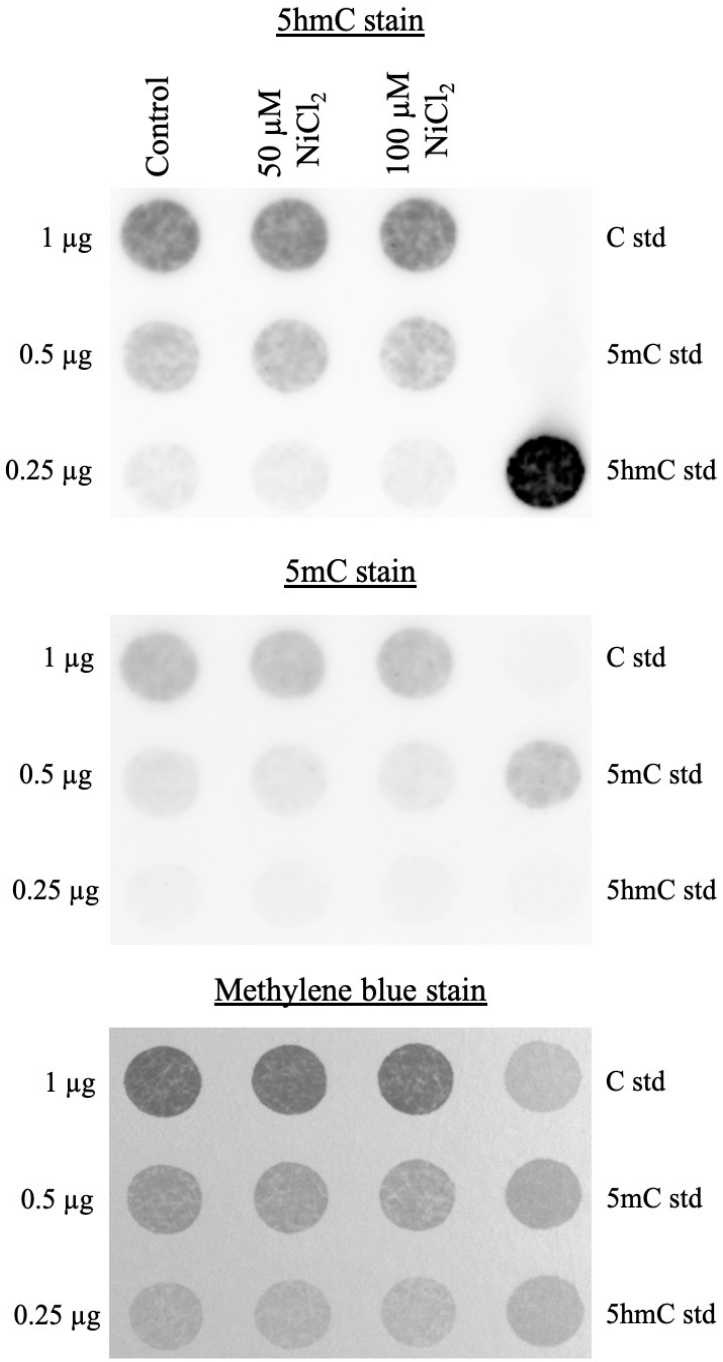
In light of the lack of expected changes in 5hmC levels despite assessing multiple different doses, no further work was conducted with $NiCl_2$ and DMαKG.

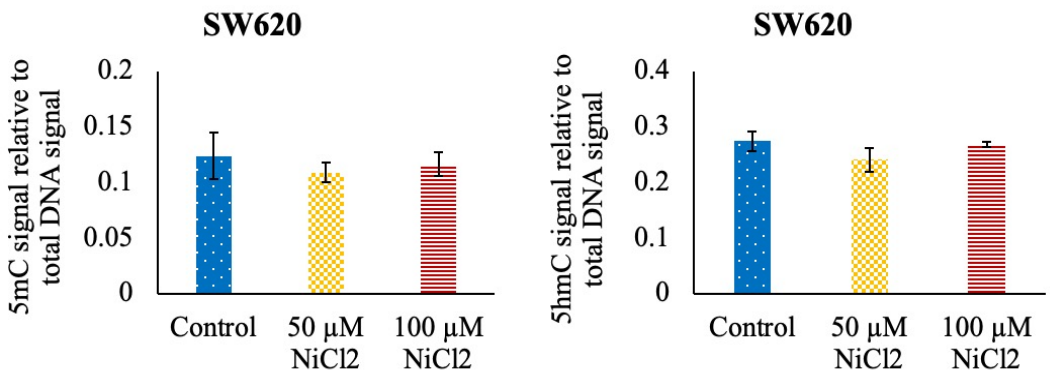
A


5hmC stain
Control
50 µM NiCl2
100 µM NiCl2
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
5mC stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
Methylene blue stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std

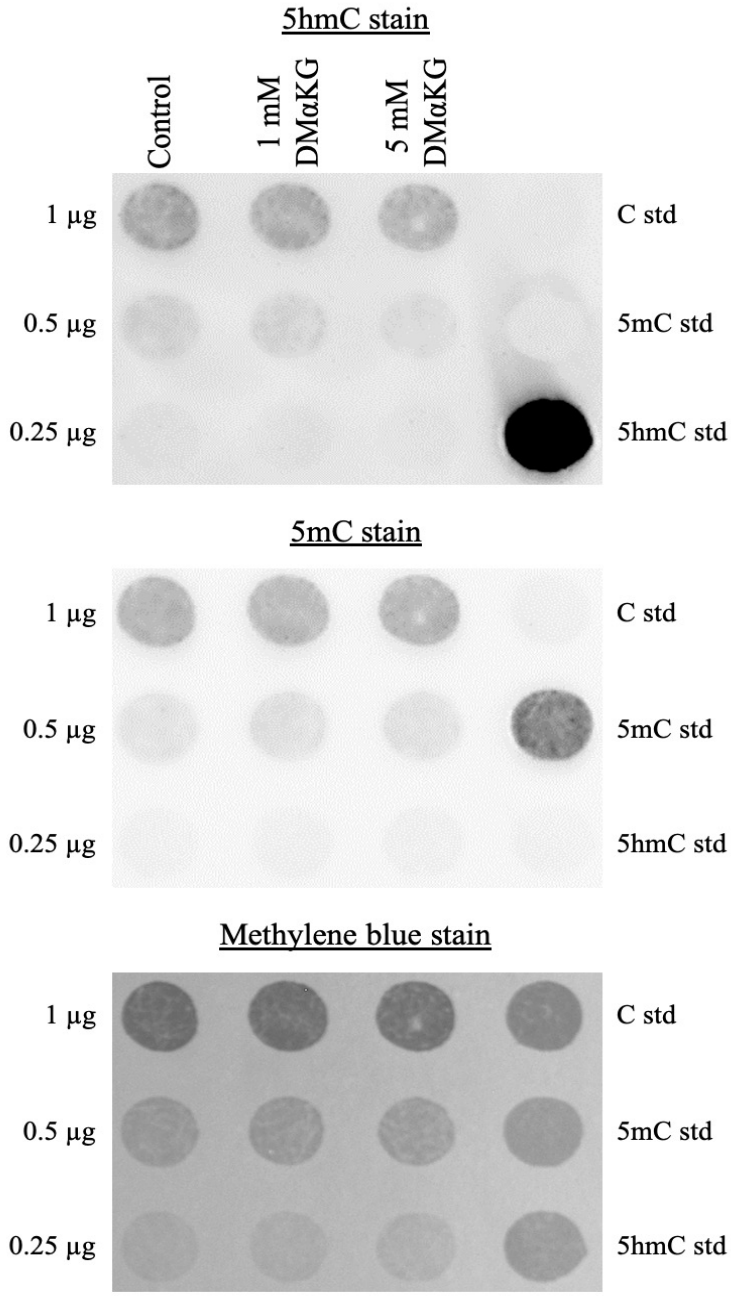

SW480
5mC signal relative to total DNA signal
0.2
0.15
0.1
0.05
0
Control
50 µM NiCl2
100 µM NiCl2
SW480
5hmC signal relative to total DNA signal
0.5
0.4
0.3
0.2
0.1
0
Control
50 µM NiCl2
100 µM NiCl2

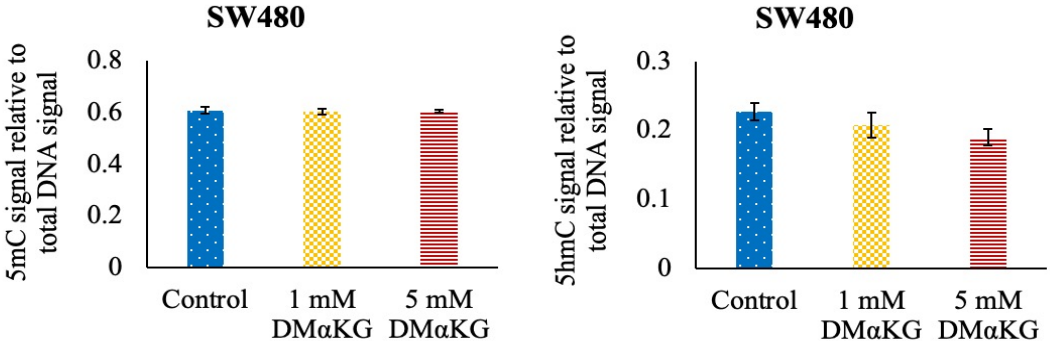
B


5hmC stain
Control
50 µM NiCl2
100 µM NiCl2
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
5mC stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
Methylene blue stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std


SW620
5mC signal relative to total DNA signal
0.2
0.15
0.1
0.05
0
Control
50 µM NiCl2
100 µM NiCl2
SW620
5hmC signal relative to total DNA signal
0.4
0.3
0.2
0.1
0
Control
50 µM NiCl2
100 µM NiCl2

C


5hmC stain
Control
1 mM DMαKG
5 mM DMαKG
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
5mC stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
Methylene blue stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std


SW480
5mC signal relative to total DNA signal
0.8
0.6
0.4
0.2
0
Control
1 mM DMαKG
5 mM DMαKG
SW480
5hmC signal relative to total DNA signal
0.3
0.2
0.1
0
Control
1 mM DMαKG
5 mM DMαKG

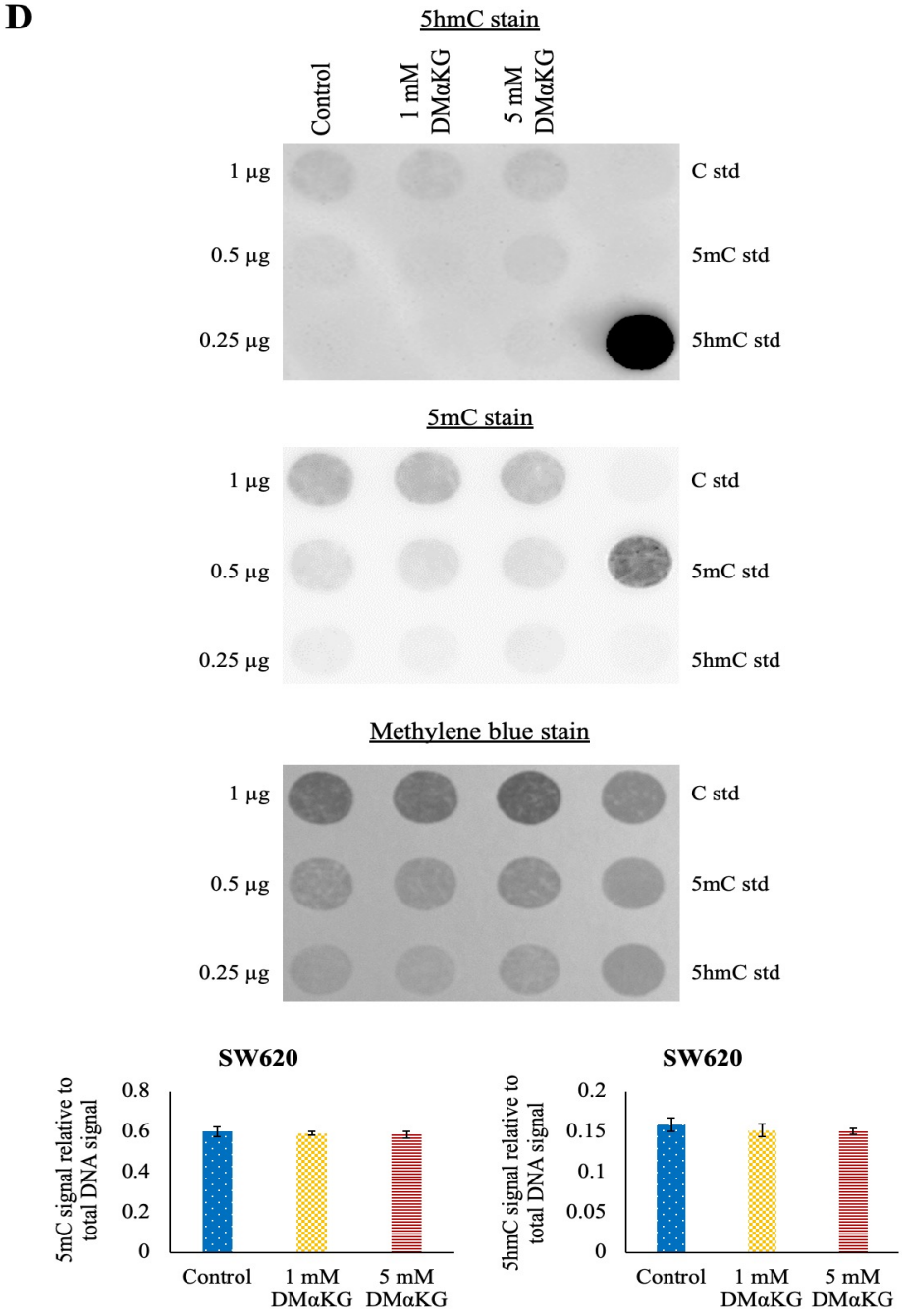


Figure 4.14 Nickel (II) chloride and dimethyl alpha-ketoglutarate treatments do not change global 5hmC and 5mC levels in SW480 and SW620 cell lines following a 48 hour treatment

(A) and (B) The representative images of 5hmC and 5mC staining in SW480 and SW620 cell line DNA samples, respectively, following a 48 hour incubation with indicated doses of nickel (II) chloride ($NiCl_2$). The errors bars represent the standard deviation of three independent replicates.

(C) and (D) The representative images of 5hmC and 5mC staining in SW480 and SW620 cell line DNA samples, respectively, following a 48 hour incubation with indicated doses of dimethyl alpha-ketoglutarate (DMαKG). The errors bars represent the standard deviation of three independent replicates.

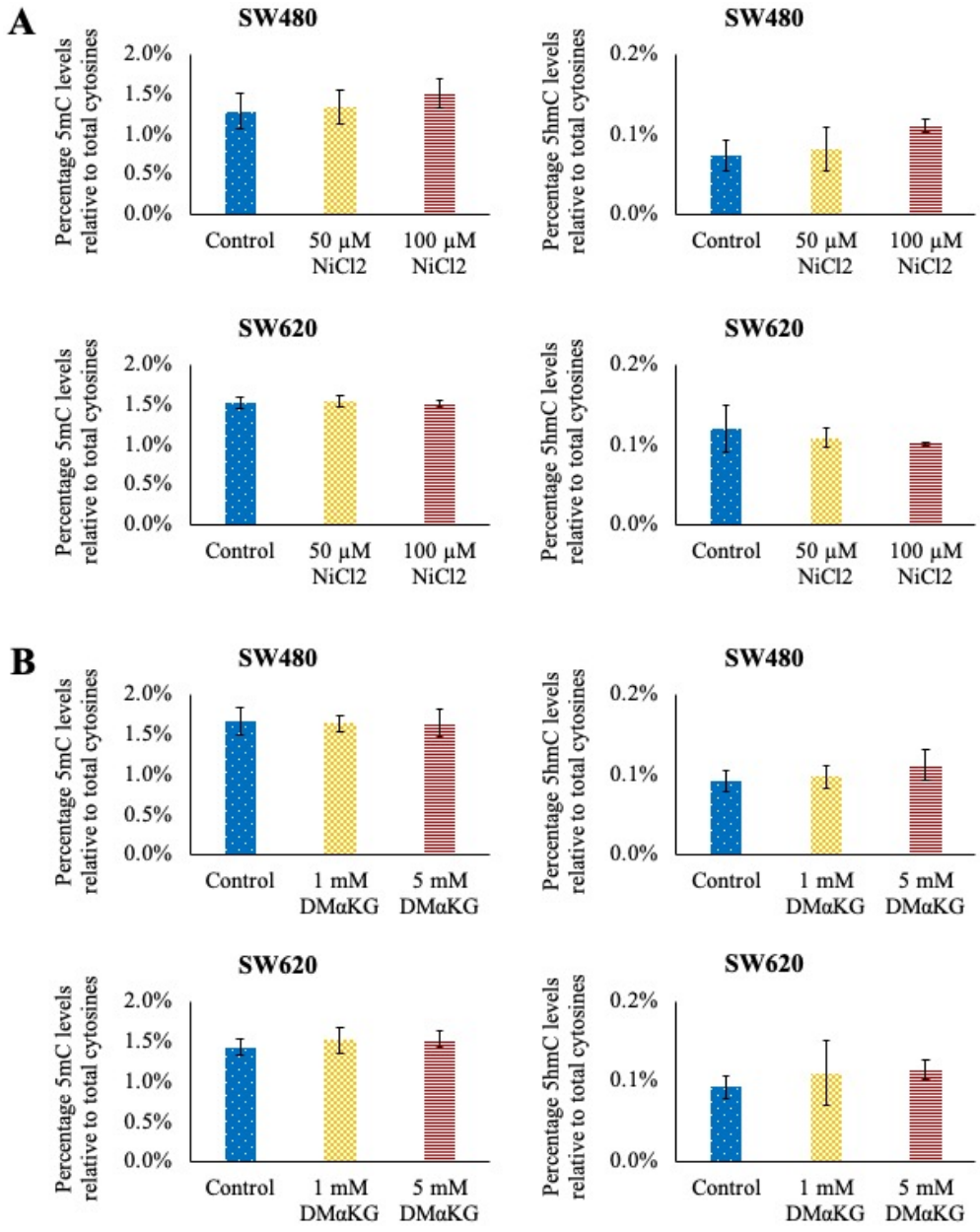
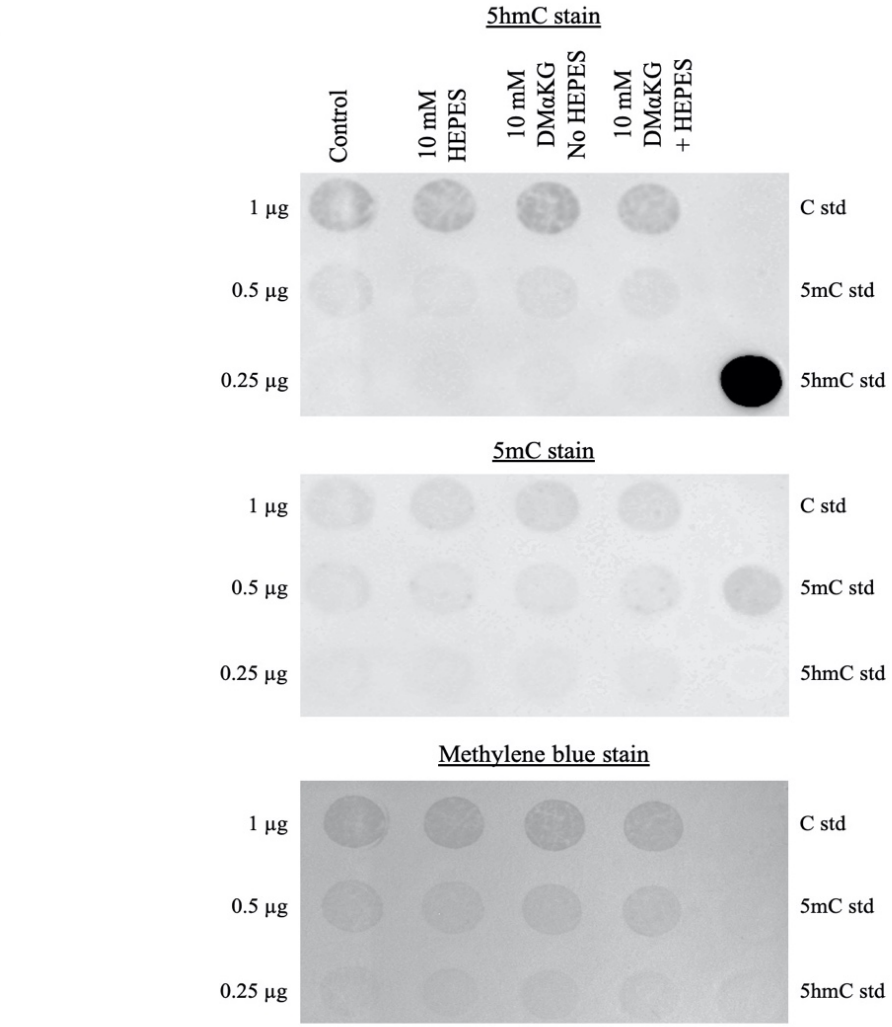


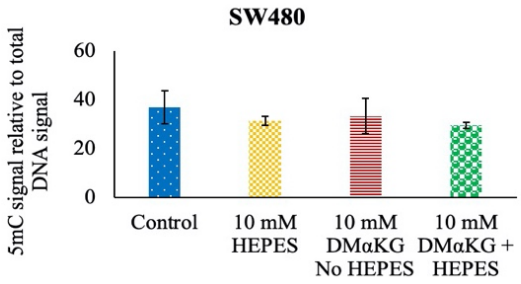
Figure 4.15 Nickel (II) chloride and dimethyl alpha-ketoglutarate treatments do not change the global 5hmC and 5mC levels in SW480 and SW620 cell lines following a 48 hour treatment

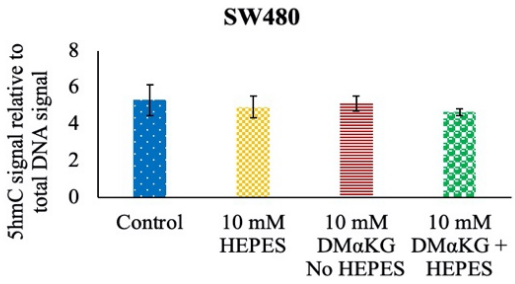
(A) The results of mass spectrometry assessment of 5hmC and 5mC levels in SW480 and SW620 cell lines following a 48 hour incubation with indicated doses of nickel (II) chloride ($NiCl_2$). The errors bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison.

(B) The results of mass spectrometry assessment of 5hmC and 5mC levels in SW480 and SW620 cell lines following a 48 hour incubation with indicated doses of dimethyl alpha-ketoglutarate (DMαKG). The errors bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison.

A


5hmC stain
Control
10 mM HEPES
10 mM DMαKG No HEPES
10 mM DMαKG + HEPES
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
5mC stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
Methylene blue stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std


SW480
5mC signal relative to total DNA signal
60
40
20
0
Control
10 mM HEPES
10 mM DMαKG No HEPES
10 mM DMαKG + HEPES


SW480
5hmC signal relative to total DNA signal
8
6
4
2
0
Control
10 mM HEPES
10 mM DMαKG No HEPES
10 mM DMαKG + HEPES

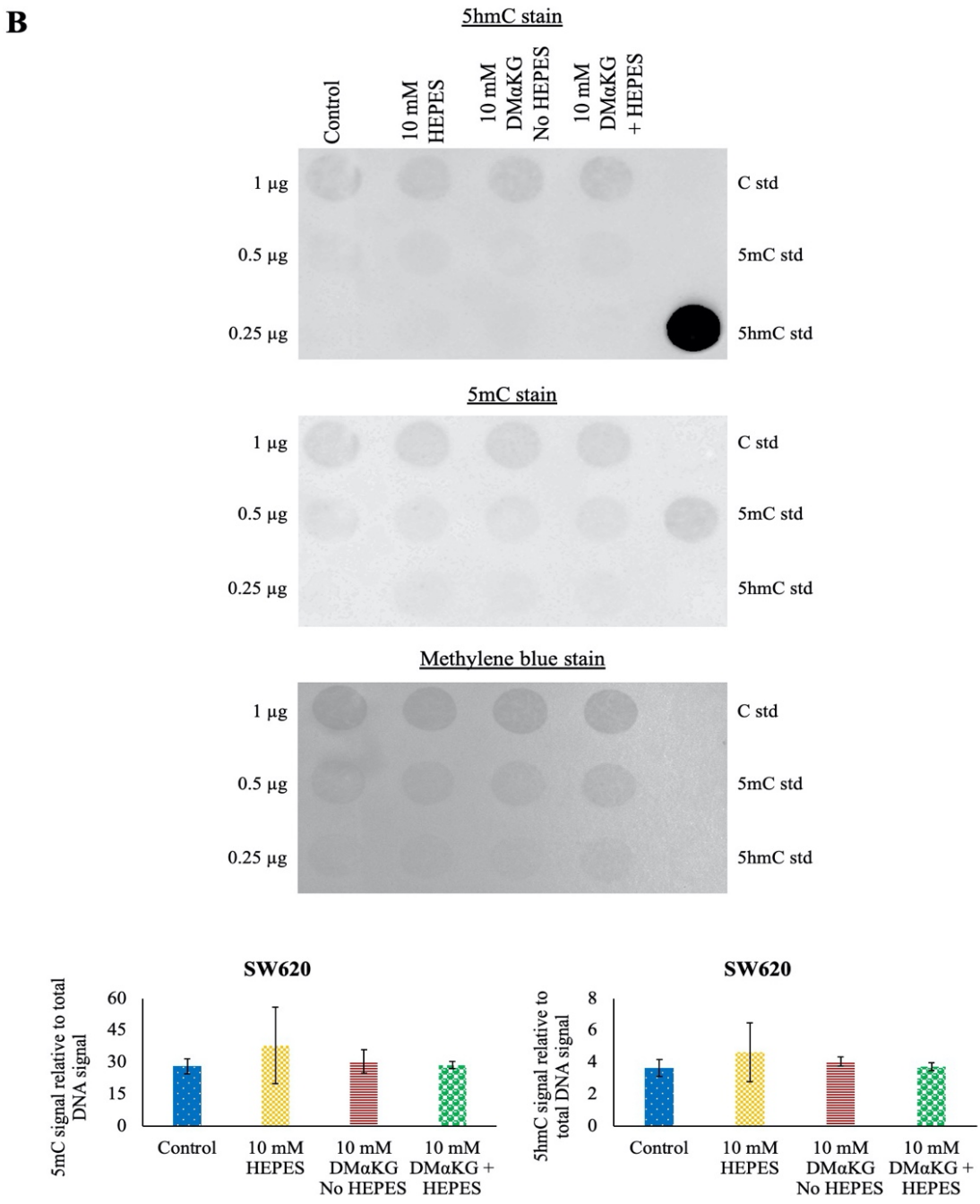


Figure 4.16 A high dose of dimethyl alpha-ketoglutarate treatment does not change global 5hmC and 5mC levels in SW480 and SW620 cell lines following a 48 hour treatment

(A) The representative images of 5hmC and 5mC staining of SW480 cell line DNA samples, following a 48 hour incubation with indicated dose of dimethyl alpha-ketoglutarate (DMαKG) with and without HEPES supplementation. The errors bars represent the standard deviation of three independent replicates.

(B) The representative images of 5hmC and 5mC staining of SW620 cell line DNA samples, following a 48 hour incubation with indicated dose of DMαKG with and without HEPES supplementation. The errors bars represent the standard deviation of three independent replicates.

4.4.4 Assessment of the impact of vitamin C on proliferation and migration rates of SW480 and SW620 cell lines

The potential anti-tumour effects of AA warrant further investigation into the stages of tumour progression it affects and the mechanisms by which this is achieved. In particular, it is critical to determine whether the effects are due to the cytotoxic properties of AA or its role in regulating the epigenome. When the effects of vitamin C on cell proliferation were assessed, the 0.1 mM dose was found to reduce the cell number of both SW480 and SW620 cell lines, but to non-significant levels (Figure 4.17 A). This AA concentration was chosen as it was found earlier to be non-toxic to the cells through the MTT assay while the higher doses greatly reduced cell viability. The introduction of the catalase enzyme allowed the use of a higher vitamin C concentration without the associated cytotoxic effect. The presence of catalase prevented any significant reduction in cell proliferation in the SW480 cell line, even at the 1 mM AA concentration (Figure 4.17 B). This suggests that the large reduction in cell viability observed through the MTT assay was a result of cytotoxicity due to the excessive hydrogen peroxide generation rather than epigenetic changes, given that addition of catalase rescued this effect. In the case of SW620 cell line, the presence of catalase did not prevent the 1 mM AA dose from significantly reducing the number of cells present in the cell culture dish well after 96 hours by 46% compared to the catalase-only control. This effect was not highly significant, however, and is in line with the general observation that the SW620 cell line appears to be more sensitive to the cytotoxic effects of vitamin C treatment.

In order to confirm these results, as well as assess the effect of AA treatment on the ability of individual cells to form independent colonies, the CFA assay was utilised and both cell lines were treated with AA with and without catalase (Figure 4.18). The colony number was significantly reduced in the SW480 cells from 71 to 14 colonies per well (Figure 4.18 A), but not the SW620 cell line (Figure 4.18 B), following treatment with 0.1 mM of AA. Conversely, the colony area was significantly reduced in the SW620 cells (by 24%), but not the SW480 cell line, compared to the untreated control. The addition of catalase rescued the reduction in the number of colonies even with the higher AA dose for both the SW480 (Figure 4.18 C) and SW620 (Figure 4.18 D) cell lines. No significant changes in colony number were observed for the SW480 cells, while the colony size did in fact increase as with both doses studied compared to both

the control and catalase only treated samples (Figure 4.18 E). The treatment with 0.1 mm dose supplemented with catalase resulted in 1.13- and 1.33-fold increase in colony area compared with untreated and catalase-only controls, respectively. The higher, 1 mM, concentration of vitamin C increased the colony area 1.13- and 1.32-fold relative to untreated and catalase-only controls, respectively. The treatment of SW620 cells with AA and catalase resulted in no significant change in neither the colony size nor area (Figure 4.18 F). This provides further evidence for the AA's effect on cell proliferation being due to hydrogen peroxide mediated cytotoxicity rather than its effect on epigenetics in SW480 and SW620 cell lines.

Next, the effects of AA treatment on cell migration were assessed in the two cell lines (Figure 4.19). The 0.1 mM dose of AA resulted in a mild reduction in the scratch diameter over the 48 hours both in terms of percentage and absolute width in SW480 cells (Figure 4.19 A). These changes were not significant. The SW620 cell line displayed an increase in the migration rate following vitamin C treatment (Figure 4.19 B). These changes also did not reach significance.

The migration rate was also unchanged by the catalase and AA treatment in both SW480 (Figure 4.19 C) and SW620 (Figure 4.19 D) cell lines. In SW480 cells, both doses of vitamin C with catalase led to a modest increase in migration rate both in terms of the absolute and percentage gap closure between the leading edge of migrating cells (Figure 4.18 E). This was also the case for the migrating SW620 cells (Figure 4.18 F), but none of these changes were found to be significant. In light of these results, it should be concluded that the epigenetic changes induced by AA treatment in the two CRC cell lines do not significantly affect proliferation, colony formation or cell migration, despite an increase in 5hmC levels.

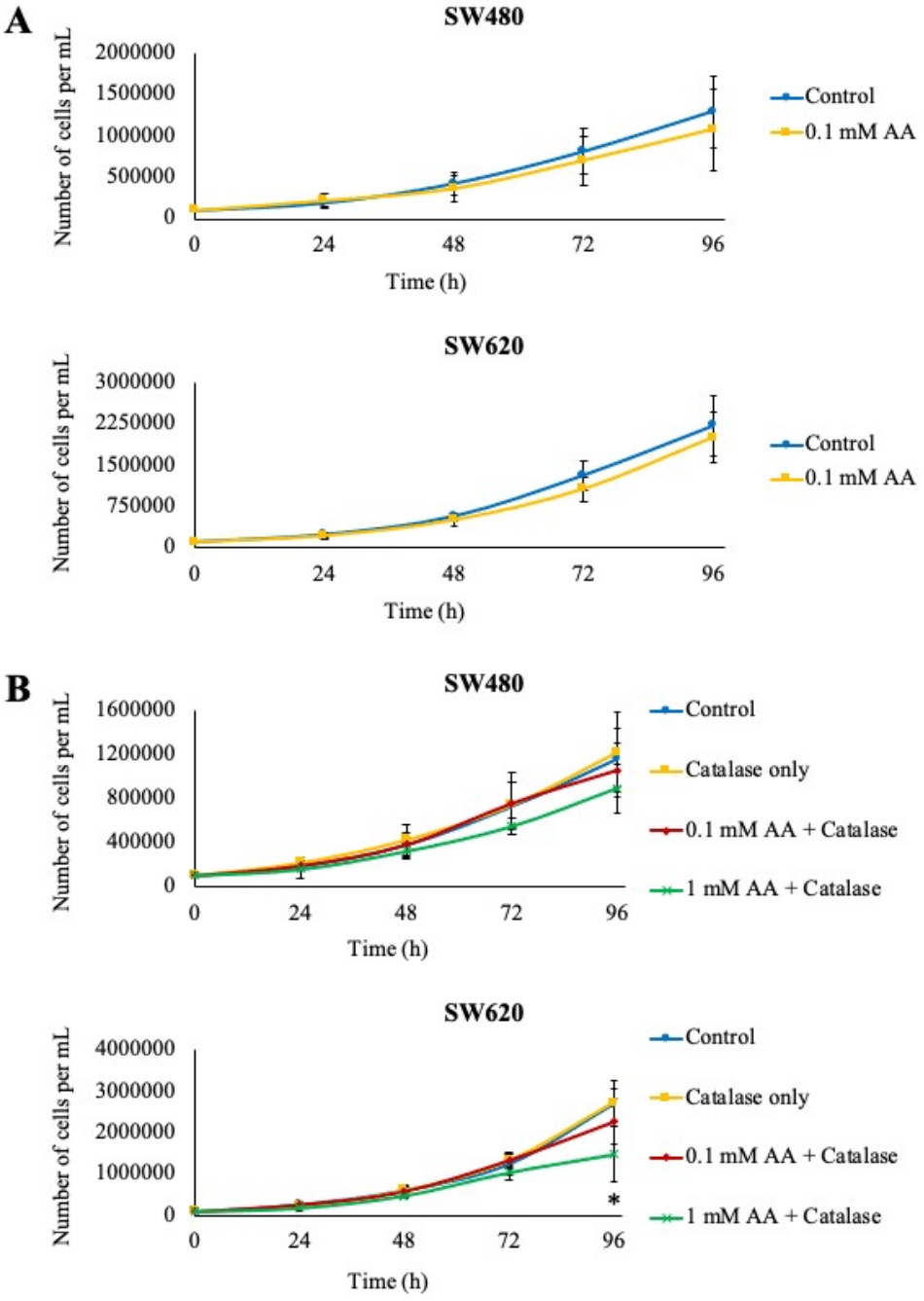
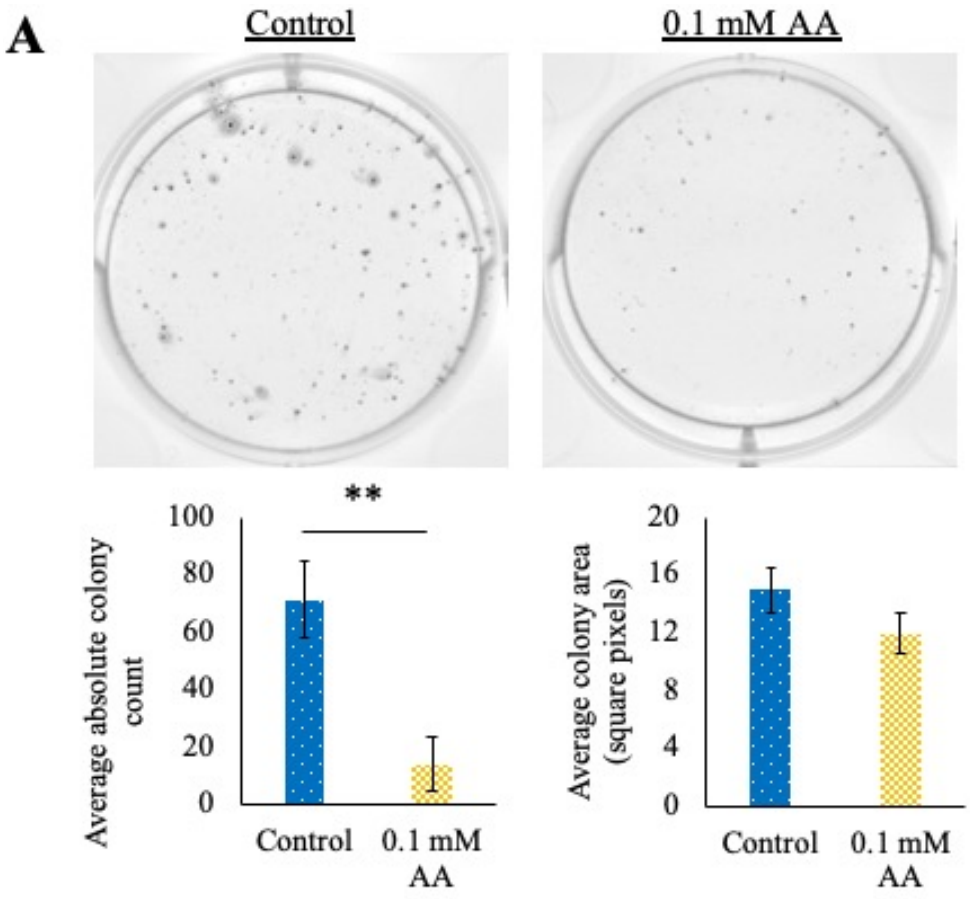
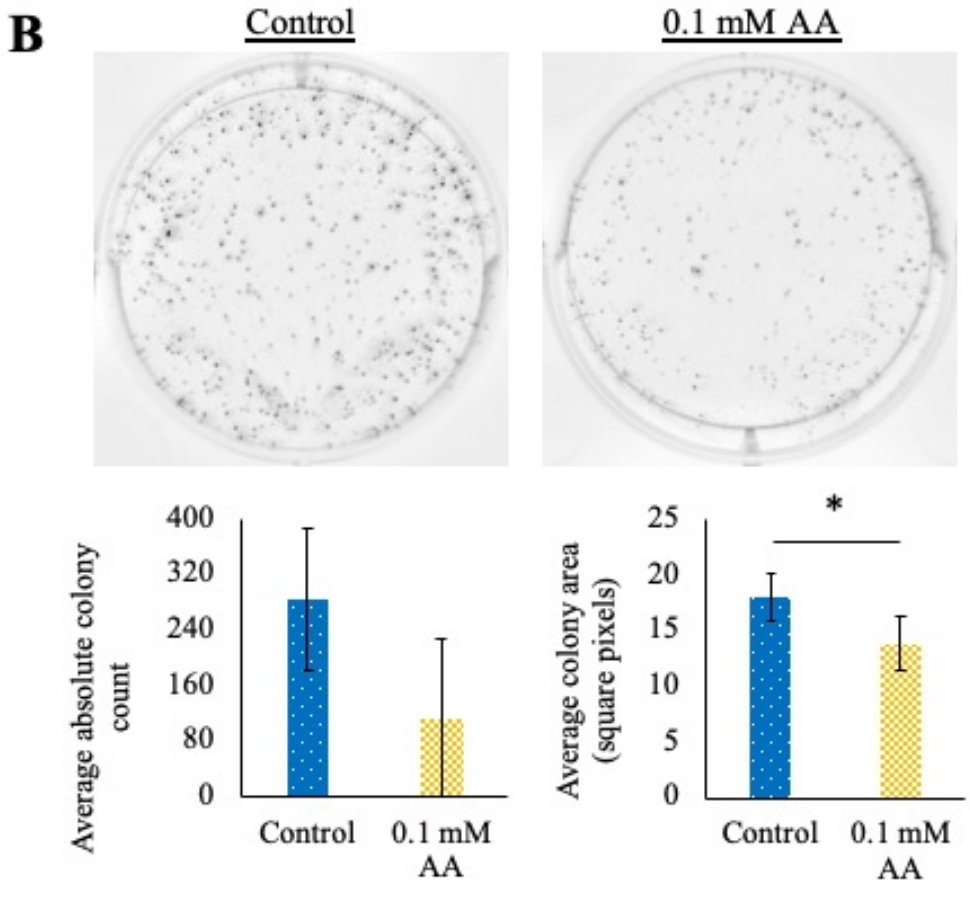


Figure 4.17 Low dose of ascorbic acid does not reduce the proliferation rate of SW480 and SW620 cells

(A) The cell counts of SW480 and SW620 treated with 0.1 mM ascorbic acid (AA) measured at 24 hour intervals for 96 hours. The error bars represent the standard deviation of six independent replicates. The significance was established using an unpaired t-test.

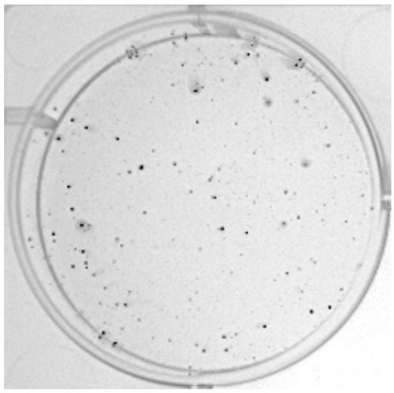
(B) The cell counts of SW480 and SW620 treated with the indicated AA doses and catalase, measured at 24 hour intervals for 96 hours. The error bars represent the standard deviation of five independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$).


A
Control
0.1 mM AA
**
Average absolute colony count
100
80
60
40
20
0
Control
0.1 mM AA
Average colony area (square pixels)
20
16
12
8
4
0
Control
0.1 mM AA

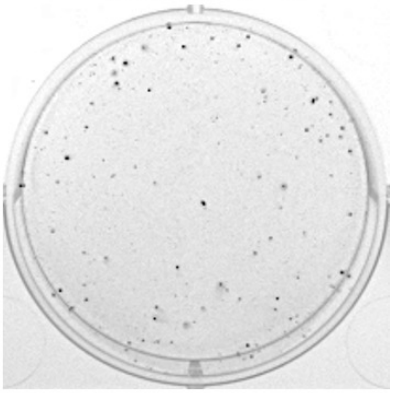

B
Control
0.1 mM AA
Average absolute colony count
400
320
240
160
80
0
Control
0.1 mM AA
*
Average colony area (square pixels)
25
20
15
10
5
0
Control
0.1 mM AA

C

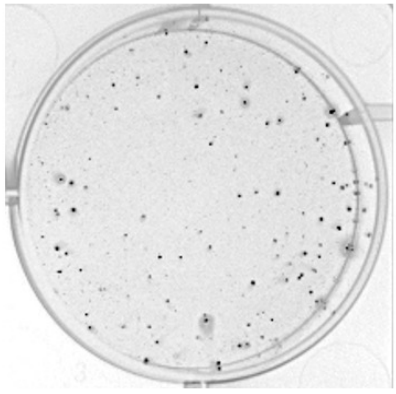
Control



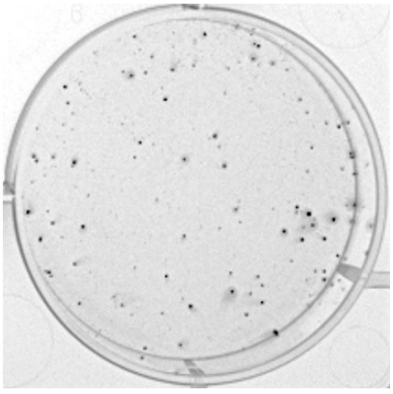
Catalase only



0.1 mM AA + catalase

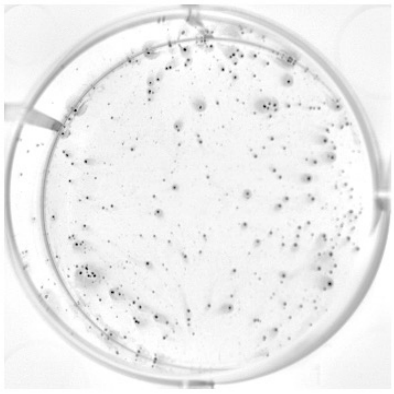


1 mM AA + catalase

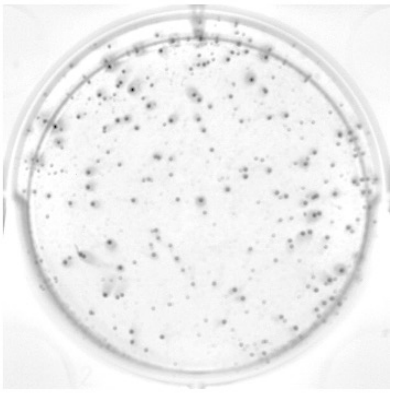


D

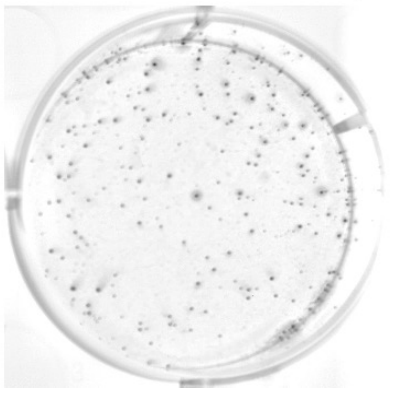
Control



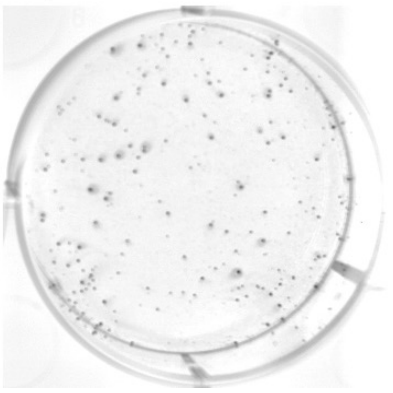
Catalase only

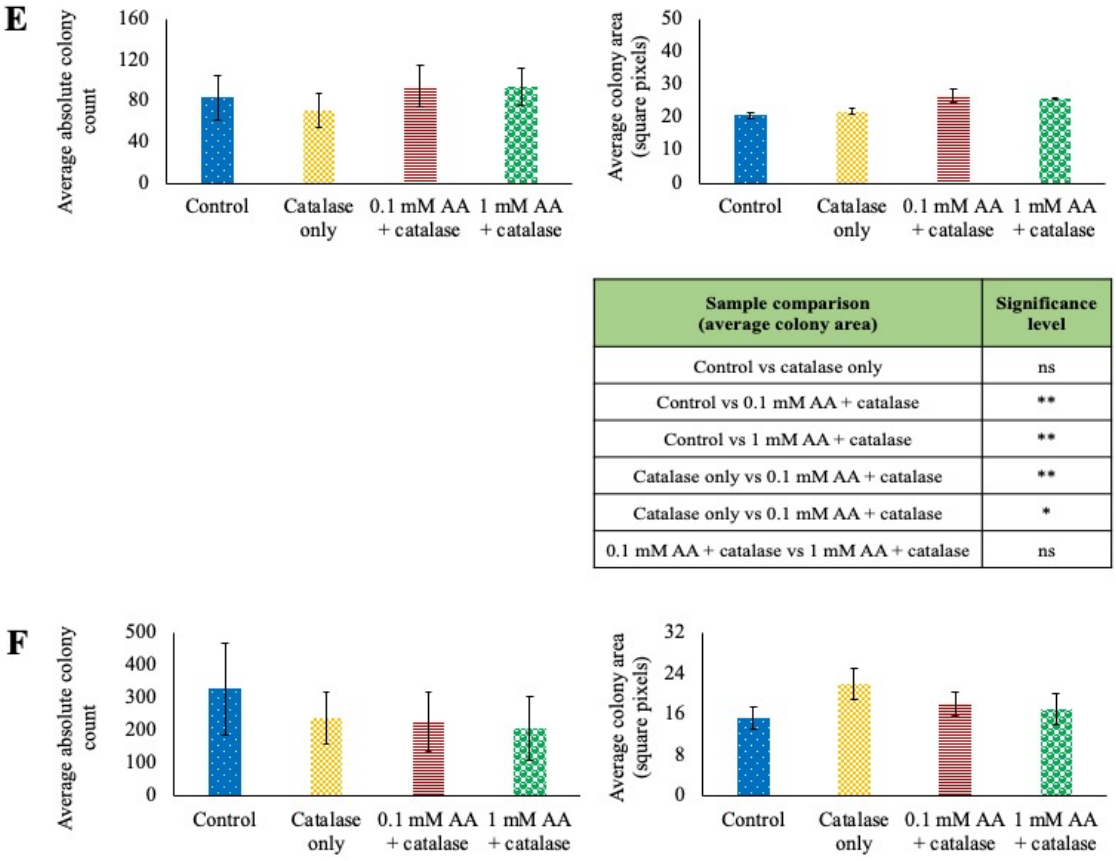


0.1 mM AA + catalase



1 mM AA + catalase





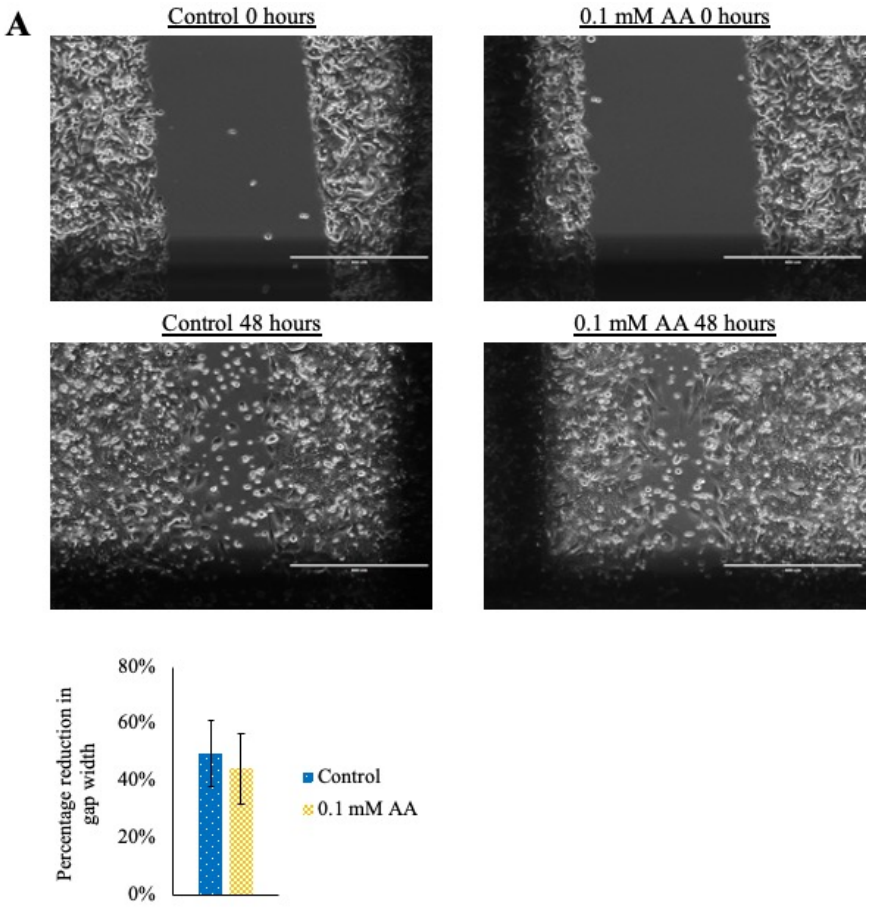
Sample comparison (average colony area)	Significance level
Control vs catalase only	ns
Control vs 0.1 mM AA + catalase	**
Control vs 1 mM AA + catalase	**
Catalase only vs 0.1 mM AA + catalase	**
Catalase only vs 0.1 mM AA + catalase	*
0.1 mM AA + catalase vs 1 mM AA + catalase	ns

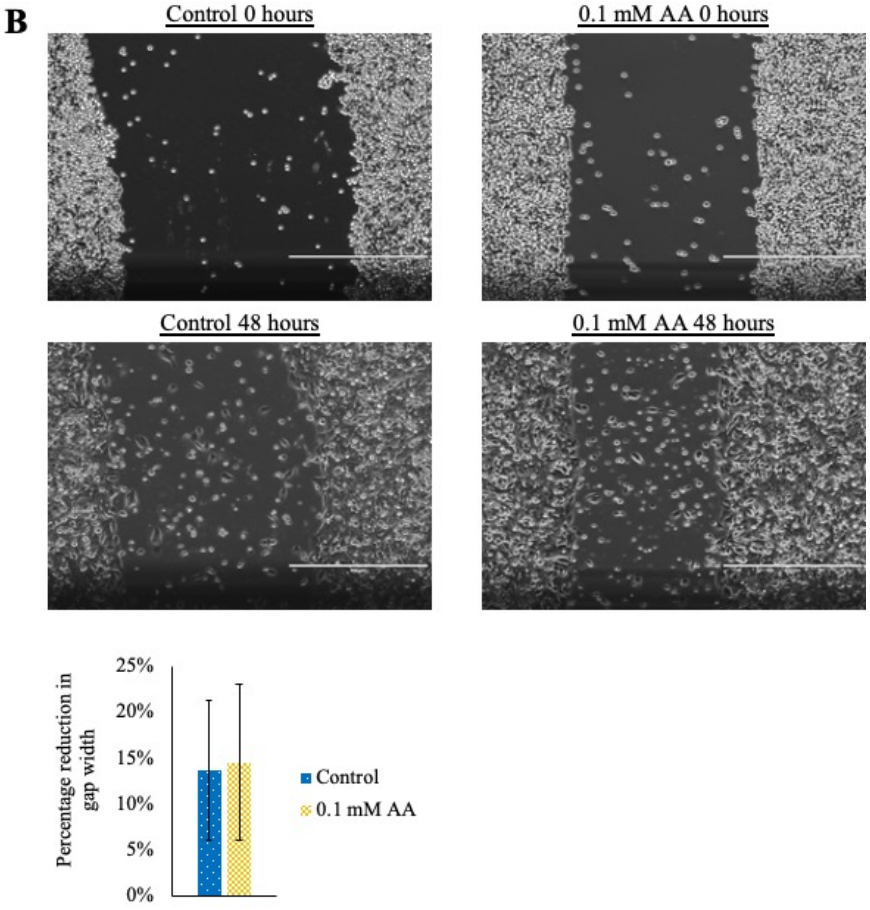
Figure 4.18 The addition of catalase rescues the ascorbic acid-induced reduction in colony formation ability of SW480 and SW620 cells

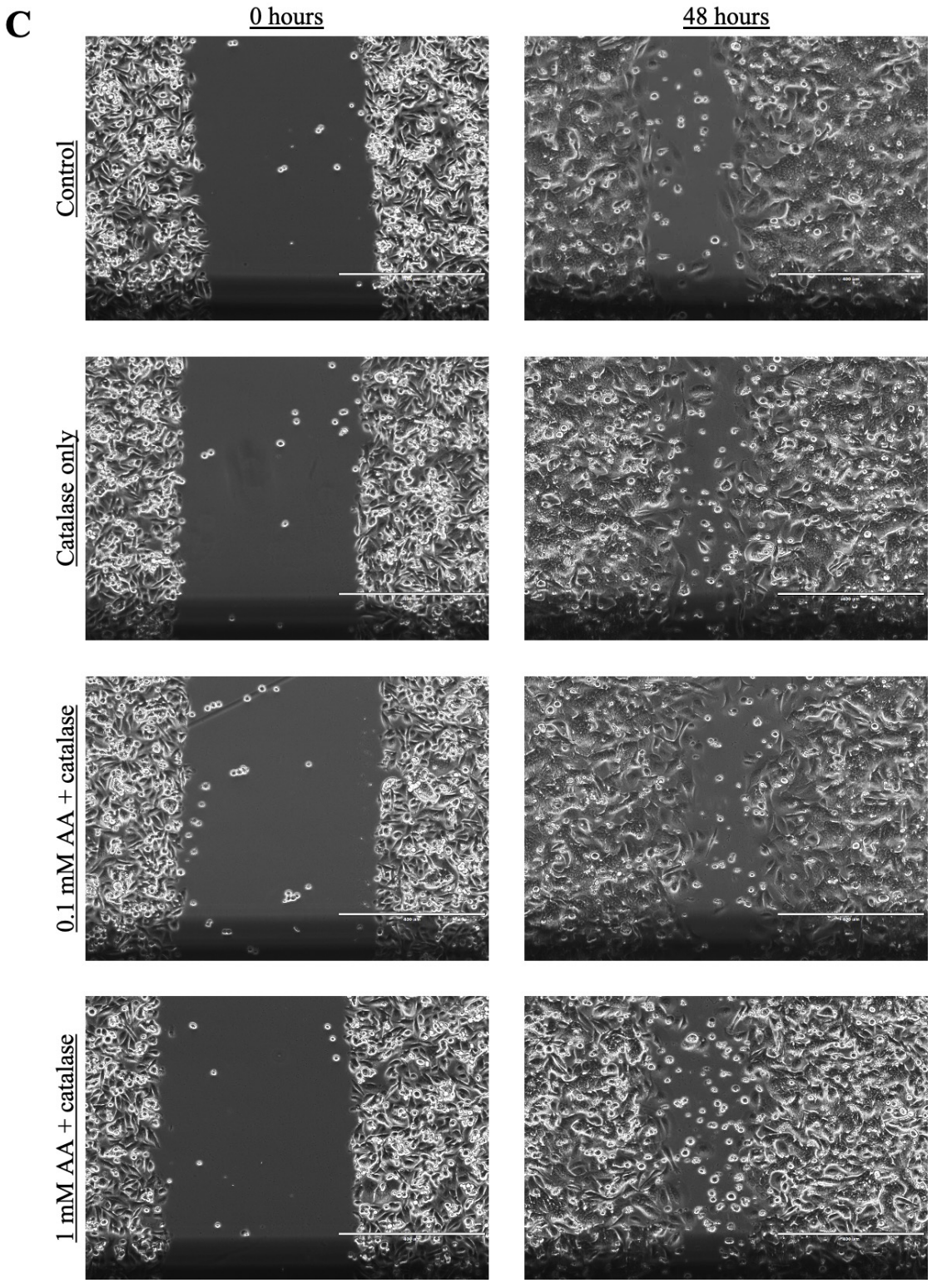
(A) and (B) The representative images and graphical representation of colony formation of SW480 and SW620 cells, respectively, following treatment with 0.1 mM of ascorbic acid (AA). The error bars represent the standard deviation of three independent replicates. The significance was established using an unpaired t-test (* $p<0.05$; ** $p<0.01$).

(C) and (D) The representative images of SW480 and SW620 cell lines', respectively, colony formation following treatment with AA and catalase.

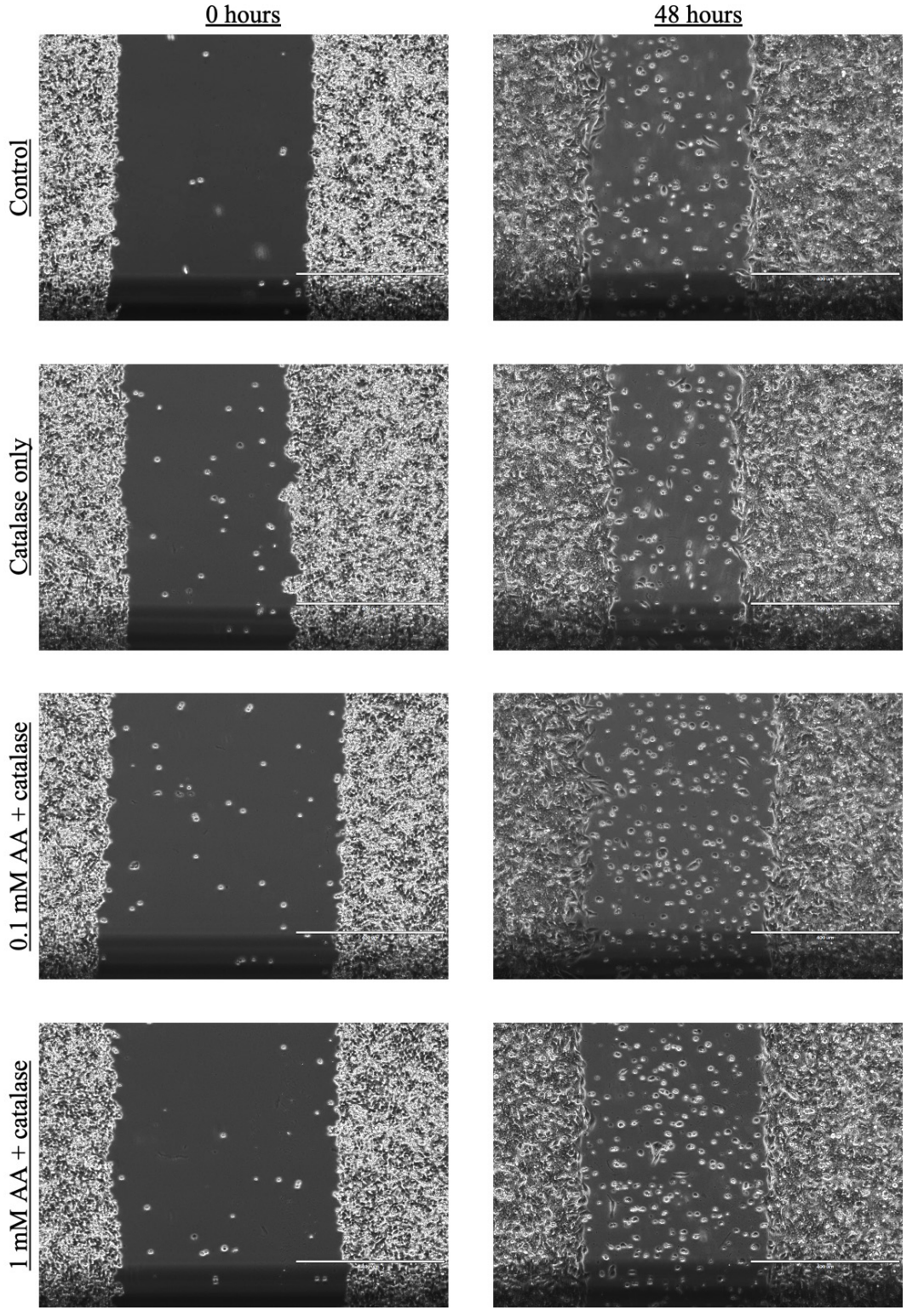
(E) and (F) The impact of AA and catalase treatment on the colony formation ability of SW480 and SW620 cells, respectively. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$; ** $p<0.01$; *** $p<0.001$). (E) The results of the statistical test are shown in the table below the graph.


A
Control 0 hours
0.1 mM AA 0 hours
Control 48 hours
0.1 mM AA 48 hours
Percentage reduction in gap width
80%
60%
40%
20%
0%
Control
0.1 mM AA


B
Control 0 hours
0.1 mM AA 0 hours
Control 48 hours
0.1 mM AA 48 hours
Percentage reduction in gap width
25%
20%
15%
10%
5%
0%
Control
0.1 mM AA


C
0 hours
48 hours
Control
Catalase only
0.1 mM AA + catalase
1 mM AA + catalase

D


0 hours
48 hours
Control
Catalase only
0.1 mM AA + catalase
1 mM AA + catalase

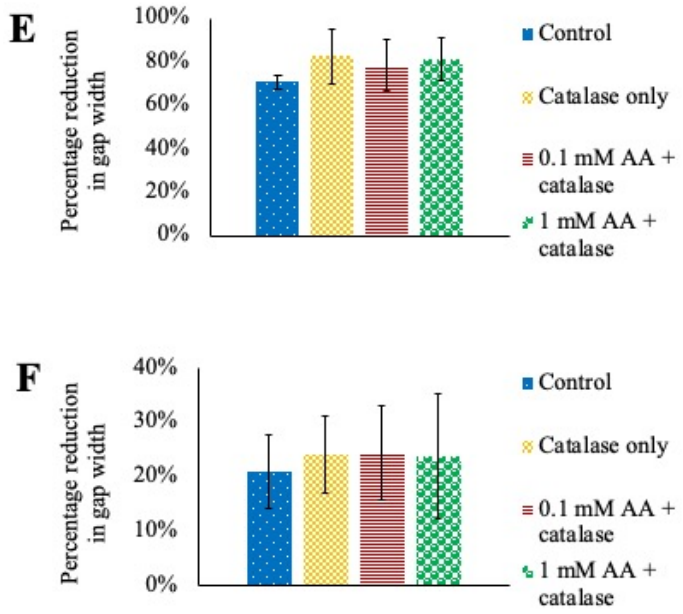


Figure 4.19 The migration of SW480 and SW620 cells in not affected by ascorbic acid treatment

(A) and (B) The representative images and results of SW480 and SW620 cell line migration, respectively, following treatment with 0.1 mM of ascorbic acid (AA). The images were taken at 0 hours and 48 hours after the scratch was made, at 10X magnification. Scale bar = 400 µm. The error bars represent the standard deviation of ten and nine independent replicates, respectively. The significance was established using an unpaired t-test.

(C) and (D) The representative images of SW480 and SW620 cell line migration, respectively, following treatment with catalase and indicated doses of AA. The images were taken at 0 hours and 48 hours after the scratch was made, at 10X magnification. Scale bar = 400 µm.

(E) and (F) The graphical representation of SW480 and SW620 cell line migration, respectively, following treatment with catalase and indicated doses of AA. The error bars represent the standard deviation of four independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison.

4.5 Discussion

This chapter describes the work performed with the aim of characterising the *in vitro* properties of the two CRC cell lines, SW480 and SW620, and the effects of compound-based 5hmC levels modulation on the *in vitro* characteristics of these cells. The results described here therefore contribute to answering the main question of this thesis: whether the TET enzymes and 5hmC contribute to the process of CRC metastasis. It provides a detailed characterisation of the two-cell line model used during this PhD project to study the metastatic characteristics *in vitro* from the epigenetic standpoint. Following that, the effects of vitamin C-based 5hmC levels modulation on these *in vitro* properties were also investigated.

The first two objectives of this chapter involved a comprehensive characterisation of the SW480 and SW620 cells lines both in terms of their *TET* expression and 5hmC levels as well as their *in vitro* proliferation, colony formation and migration. *TET1* expression was shown to be elevated in the metastatic SW620 cells while the opposite trend was observed for *TET2* levels, which were significantly higher in the primary SW480 cells. No changes were seen in *TET3* expression. These changes were also confirmed by assessing the absolute expression of each *TET* gene in the two cell lines. TET2 and TET3 protein levels were found to closely follow the transcript expression analysis results: TET2 was found to be significantly elevated in the primary cell line while TET3 remained unchanged. TET1 levels on the other hand were found to be the same in the two cell lines, which is in stark contrast to the qRT-PCR data. This discrepancy could be explained by poor specificity of the anti-TET1 antibody used given the numerous non-specific bands present in the blots as shown in Figure 4.3. Additionally, the TET1 protein could be differently regulated by post-translational modifications in the two cell lines which would result in the same protein levels despite differences in *TET1* transcript content.

The upregulation of *TET1* in SW620 cells as compared to the SW480 cell line was also observed by two independent groups (Kai et al., 2016; Tian et al., 2017). On the other hand, Huang and colleagues reported little change in *TET1* levels between the two cell lines, although they observed the same expression pattern for *TET2* and *TET3* as reported here (Huang et al., 2016). Wu and co-authors reported a higher expression of all three *TET* transcripts in the metastatic SW620 cells, with the most pronounced

difference observed in the *TET3* levels (Wu et al., 2016). These discrepancies are likely due to different reference genes used and differences in the primer efficiencies and exons targeted by the *TET* primers. The geNORM technique was utilised in this project to find the most stable reference gene in the SW480 and SW620 cell lines, which was found to be *ACTB*. This further strengthens the results described in this chapter.

Huang and colleagues reported undetectable TET1 protein levels in the SW620 cell line (Huang et al., 2016). Another publication found all three TET enzymes to be present at similar levels in the two CRC cell lines with TET1 slightly more abundant in SW620 cells and TET3 more prevalent in SW480 cells (Wu et al., 2016). The two articles and the work presented in this chapter used different antibodies to detect the levels of the three TET proteins which could explain the differences in results observed (Huang et al., 2016; Wu et al., 2016). The direct comparison of the 5hmC levels could not be found in the literature. It is likely that the higher global abundance of this cytosine modification in SW480 cells is due to the higher overall levels of *TET* transcripts as determined by the absolute quantification qRT-PCR.

The cell proliferation rate was higher in the metastatic SW620 cell line as compared to its primary counterpart, the SW480 cell line. The same result was previously published in the literature and could reflect the fact that the SW620 cells had to increase their proliferation rate in order to establish the secondary niche in the lymph node of the patient (Schønberg et al., 2006; Lei et al., 2011) . On the other hand, the proliferation of the SW480 cells might have been limited in the primary tumour by its size and the blood supply. The opposite pattern was seen for migration, where the primary SW480 cell line had a significantly higher migration rate than the SW620 cells. This result is also in line with the previously published work (Kubens and Zänker, 1998; Lei et al., 2011; Slater et al., 2018) and could indicate that the SW480 cells isolated from the primary tumour are progressing in the carcinogenesis process toward a metastatic phenotype. On the other hand, the SW620 cells having already metastasised *in vivo* are more likely to prioritise growth and survival over migration.

It is important to note that the SW480 and SW620 represent a limited and very simplistic model of colorectal cancer metastasis. The two cell lines have been in culture for over four decades and likely lost numerous properties and characteristics they had before they were taken from the patient. The SW620 cell line was obtained from a

biopsy taken from a lymph node metastasis and thus does not represent a good tool for studying colorectal cancer metastasis to the liver, which is the main focus of this PhD project. Lastly, the two cell lines represent a two distinct time points in the tumorigenesis and metastatic cascade and cannot hence be used to investigate which steps of these processes could be influenced by the TET enzymes and pathological alterations in global 5hmC levels. Nevertheless, SW480 and SW620 cell lines form an easy to use, cheap and time-efficient *in vitro* model that can be used for an initial assessment of the likely roles of the TETs in CRC progression. The results obtained from this investigation, however, should be interpreted with the abovementioned limitations in mind.

The second set of objectives for this part of the PhD project was to assess the impact of treatment with compounds modifying 5hmC levels on the *in vitro* proliferation, colony formation and migration of SW480 and SW620 cell lines. The best known and studied of the molecules chosen is AA. The potential anti-cancer properties of vitamin C have been studied for over half a century. Despite the initial scepticism, the evidence for its efficacy in cancer treatment has been accumulating, particularly as an adjuvant to chemotherapy. The *in vitro* and *in vivo* tumour growth suppression caused by AA has largely been attributed to generation of hydrogen peroxide and reactive oxygen species (Park et al., 2004; Chen et al., 2008). This observation is supported by the rescue of cell viability observed for both SW480 and SW620 cells when the physiological AA concentration (1 mM) (Lykkesfeldt and Tveden-Nyborg, 2019) was supplemented with catalase. Additionally, catalase was capable of preventing the reduction in proliferation observed for all doses of AA administered to the two cell lines. This further reinforced the hypothesis that the reduction in proliferative capacity of SW480 and SW620 cells following AA treatment was due to cytotoxicity associated with excessive hydrogen peroxide production rather than its effects on the activity of TET enzymes.

The low dose (0.1 mM) of AA was not found to affect the migration of the SW480 and SW620 cell lines. Similarly, no changes were observed in the migration of either cell line when the 1 mM vitamin C and catalase were added to cell culture medium. It is tempting to speculate that the 1 mM dose of AA used on its own would have resulted in a significant reduction in migration of the two cell lines. Indeed, high doses of AA were found to reduce cell migration both *in vitro* and *in vivo* in different cancer types (Polireddy et al., 2017; Gan et al., 2019). However, high doses of AA on its own were

not tested in this project as they were found to result in death of the majority of cells in culture. Additionally, the focus of this study was on the epigenetic changes, and more specifically the increase in global 5hmC levels, induced by vitamin C rather than its cytotoxic effects in cancer cells.

The AA treatment, with and without catalase, resulted in no changes in the expression of the TET enzymes, which is in line with the evidence that vitamin C affects TET enzyme activity rather than the expression of the genes. Indeed, the same dose of AA as the one used in this study (1 mM) was recently shown to have no effect on the expression of the TET enzymes in the CRC HCT116 cell line (Gerecke et al., 2020).

Perhaps the most unexpected result obtained during this study was a small, non-significant upregulation of global 5hmC levels following the AA treatment. The vitamin C on its own led to a modest increase in 5hmC levels while the global 5mC abundance was unchanged despite studies suggesting that AA treatment of cancer cell lines reduces their 5mC content (Cimmino et al., 2018; Ge et al., 2018). This could reflect the comparatively modest increase in 5hmC levels observed in this study. Similarly, the higher doses of AA supplemented with catalase led to a modest, non-significant increase in 5hmC levels as measured by mass spectrometry. This is in contrast to results of multiple previously published studies where similar doses of vitamin C resulted in a few-fold elevation of global 5hmC content in numerous cancer cell line types (Ge et al., 2018; Peng et al., 2018; Gerecke et al., 2020). This discrepancy could be explained by different methodologies used to measure the 5hmC levels. Out of the studies referenced above, only Gerecke and colleagues utilised the gold standard measurement of 5hmC levels which is liquid chromatography coupled to mass spectrometry. The other two studies only used the dot blot staining, which is semi quantitative and prone to measurement error due to unequal DNA loading and potential for antibody cross-reactivity. The potential for discrepancy between the two methods is best illustrated by the results shown in Figures 4.10 C and D and Figure 11 B. While the dot blots results indicated a between 3.5- and 5.1-fold upregulation of 5hmC levels in the AA and catalase treated samples, the mass spectrometry data showed a more modest, approximately 1.5 to 1.8-fold increase in global 5hmC abundance which was not found to be significant. When such contradictions arise, the mass spectrometry results should be accepted as more accurate given the superior sensitivity and fully quantitative nature of this technique. Additionally, the different formulations of AA used by the different

research groups and in this project may exhibit varying stability in the cell culture media and thus lead to different outcomes in terms of increasing the activity of the TET enzymes (Ge et al., 2018; Peng et al., 2018). Lastly, the high proliferation rate of the SW480 and SW620 cell lines may limit the potential for an increase in 5hmC levels following AA treatment due to insufficient time available within the cell cycle to re-establish the new patterns of 5hmC distribution caused by vitamin C.

An intriguing concept was recently proposed by Cho and colleagues regarding a differential effect of low and high AA doses on CRC cells (Cho et al., 2018). The low doses, such as those achieved with oral administration, were proposed to generate insufficient hydrogen peroxide and reactive oxygen species to kill cancer cells and hence to induce a hormetic response which leads to an increased tumour growth. On the other hand, high doses, such as the ones obtained with intravenous infusions, were proposed to create sufficient oxidative damage to kill cancer cells. The authors were indeed able to show the hormetic response taking place in CRC cell lines expressing low levels of the SVCT-2 vitamin C transporter. This theory could provide an explanation for the different results obtained by Linus Pauling and collaborators who administered a high IV dose to the patients and the studies subsequently released by the Mayo clinic which relied on oral route of administration and consequently achieved a lower concentration in the study subjects.

In addition to the findings described above, previously published research does support 5hmC and TET enzymes-dependent effects of AA treatment of cancer cells. Indeed, it was found to reduce cell proliferation and tumour invasiveness in lymphoma and melanoma cell lines, respectively, in a hydrogen peroxide-independent manner while raising the global 5hmC levels (Gustafson et al., 2015; Shenoy et al., 2017). This mechanism may help to explain the anti-metastatic effects observed in the PANC-1 pancreatic cell line *in vitro* as well as the 4T1 breast and B16FO melanoma cell lines *in vivo* (Cha et al., 2013; Polireddy et al., 2017). In summary, both the cytotoxic effects of vitamin C as well as its effects on cancer cells through epigenetic pathways clearly warrant further investigation based on the data outlined above. Given that AA is an essential nutrient in humans and many cancer patients suffer from vitamin C deficiency (scurvy) (Fain et al., 1998; Klimant et al., 2018) combined with the established safety of the high intravenous doses in patients and the cytotoxic effects that appear to be specific to cancer cells, the use of AA is an exciting prospect in cancer therapy once again.

The incubation of SW480 and SW620 cells with $NiCl_2$ and DMαKG resulted in little changes in transcript levels of the three TET enzymes, with *TET1* elevated slightly under most of the treatment conditions. While the literature on the subject is scarce, this result was expected given that these compounds are proposed to only modulate the TET activity rather than expression of these enzymes (Yang et al., 2016). In particular, no data is available on the effect on $NiCl_2$ treatment on the expression of *TET1*, *TET2* and *TET3* genes in culture and hence this represents a novel contribution to our understanding of the effect of this compound on the epigenetic machinery in cancer cell lines.

The treatments with the two compounds did not result in the expected changes in global 5hmC levels in SW480 and SW620 cell lines. The results of both the dot blot and mass spectrometry analyses indicate that the abundance of 5mC and 5hmC in both cell lines remained constant under all concentrations used. In fact, even the trial of a much higher dose of DMαKG (10 mM) did not lead to the expected upregulation of 5hmC levels. In light of this fact, further work was not conducted with these compounds. It is not clear why these treatment did not produce the expected outcome in terms of changing the global 5hmC content, despite their previous successful use in the literature (Yang et al., 2016; Yin et al., 2018). The SW480 and SW620 cells might already contain sufficiently high concentration of alpha-ketoglutarate and iron (II) to ensure the optimal function of the TET enzymes. Additionally, shorter or longer incubation times with the two compounds might be required in order to see the expected changes in 5hmC levels due to their stability and turnover in culture.

This study represents a solid starting point for further investigation of how compounds modulating 5hmC levels affect *in vitro* cancer properties. The main limitation of the study is the use of the three compounds for only one, 48 hour, time period (aside from the proliferation and colony formation assays). A time course experiment assessing the TET activity with multiple time points ranging from hours up to a week could provide the insight into what treatment period is optimal for achieving the greatest effect on function of the TET enzymes. This would then better inform the time period for which the functional assays should be conducted and how often should the cells be re-supplemented with the compounds. An interesting example of the importance of treatment period was reported by Shenoy and colleagues where lymphoma cells showed a gradual reduction in cell viability following a 72 hour treatment period with a high

dose (5 mM) vitamin C treatment supplemented with catalase (Shenoy et al., 2017). This indicates tumour suppression via hydrogen peroxide-independent, possibly 5hmC-depenendent, mechanism. Lastly, as this project is focused on the possible epigenetic contribution to CRC metastasis, the use of the three compounds described in this chapter in conjunction with a cell invasion assay, such as a Boyden chamber containing extracellular matrix-like mesh, would be the next experiment that should be conducted as a follow up to this study. This would allow one to more closely recapitulate the tumour microenvironment and the effect of varying doses of the 5hmC-modifying compounds on cancer migration and invasion processes.

Chapter 5

Genetic manipulation of TET enzymes in an *in vitro* colorectal cancer metastasis model

5.1 Chapter summary

The aim of this investigation was to elucidate the effects of genetic modulation of TET expression on the proliferation and migration of SW480 and SW620 colorectal cancer cell lines. This work was undertaken in order to complement the results obtained with compound-mediated modifications of 5hmC levels as described in Chapter 4.

CRISPR/Cas9 TET knockouts of SW480 and SW620 cell lines generated by Horizon Discovery were characterised and found to have reduced transcription of the *TET* genes. This reduction was confirmed through protein levels analysis, in particular in the primary cell line. Unexpectedly, global 5hmC content persisted in the DNA albeit at reduced levels. Following these findings, the cells were considered as *TET* knockdowns (KDs) rather than complete knockouts as initially thought.

Reduction in TET expression was found to result in a lower proliferation rate of the TET KD cell lines. The TET knockdowns had different effects in the two colorectal cancer cell lines: they resulted in higher migration rate of the primary SW480 cell line and they reduced the migration of metastatic SW620 cells. The overexpression of TET1 and TET3 enzymes in the two TKD SW480 cell lines resulted in a partial rescue of the migratory phenotype while little effect of individual TET overexpression on migration rate was observed in the TKD SW620 cell line.

Taken together, these results provide an in-depth description of the effects of genetic modulation of TET enzymes levels on the *in vitro* proliferation and migration of the SW480 and SW620 cell lines and provide evidence for opposite effects of these enzymes on tumour characteristics based on the stage of colorectal cancer progression.

5.2 Introduction

The reduction in global 5hmC levels, a hallmark of multiple tumour types, was found in CRC biopsies (Uribe-Lewis et al., 2015; Gilat et al., 2017; Tian et al., 2017). Intriguingly, there are conflicting reports regarding the expression of the *TET* genes in CRC tumour biopsies. Uribe-Lewis and colleagues found that the mRNA levels of all three *TET* genes were unchanged in adenomas and adenocarcinomas of the colon compared to normal tissue (Uribe-Lewis et al., 2015). A different study reported that *TET1* levels were reduced in half of the analysed CRC samples (Kudo et al., 2012). Rawłuszko-Wieczorek and colleagues found that the expression of *TET1*, *TET2* and *TET3* were all reduced in CRC patient biopsies compared to surrounding colonic tissue (Rawłuszko-Wieczorek et al., 2015). They also found a positive correlation between high *TET2* expression and overall survival in CRC sufferers.

Two independent groups reported the reduction in *TET1* levels in CRC patient samples and confirmed that a *TET1* knockdown leads to increased *in vitro* proliferation (Neri et al., 2015; Tian et al., 2017). The preliminary data indicating a potential role of the TETs in CRC metastasis (section 4.2, Figure 4.1) is supported by the findings of Puig and collaborators (Puig et al., 2018). They demonstrated an increase in proportion of high-5hmC levels biopsies in the liver metastases compared to matched primary CRC patient samples. High TET2 and 5hmC levels were also negatively correlated with disease-free survival of CRC sufferers following chemotherapy treatment. In light of the evidence described above, the role of the TET enzymes and 5hmC in CRC progression warrants further investigation.

In order to investigate the role of the TET enzymes in the SW480 and SW620 cell line model of CRC metastasis, knockout cell lines were generated using the clustered regularly interspaced palindromic repeats/CRISPR-associated 9 (CRISPR/Cas9) technology by Horizon Discovery (see Supplementary Figure 6 for details). The cell lines created include a *TET2* single knockout (SKO), *TET2* and *TET3* double knockout (DKO) and *TET1*, *TET2* and *TET3* triple knockout (TKO) in each of the SW480 and SW620 cell lines. Two different TKO clones were generated in SW480 cells, termed TKO 1 and TKO 2. This chapter presents the characterisation of all the above-mentioned cell lines and the effect of TET knockouts on *in vitro* cell proliferation, colony formation and migration.

5.3 Experimental objectives

Aim: to examine the effects of CRISPR/Cas9-induced mutations in genes encoding all three TET enzymes on *in vitro* proliferation, colony formation and migration of the SW480 and SW620 colorectal cancer cell lines.

Objectives:

1. Confirm the CRISPR/Cas9 -induced knockouts at mRNA and protein levels and assess their effect on the global 5hmC content in SW480 and SW620 cell lines
2. Establish the impact of the single, double and triple *TET* knockouts on cell proliferation, colony formation and migration in the SW480 and SW620 cells
3. Examine the effect of the TET rescue through exogenous overexpression in the triple knockout SW480 and SW620 cell lines

5.4 Results

5.4.1 Confirmation of the CRISPR/Cas9 -induced knockouts at mRNA and protein levels and assessment of their effect on the global 5hmC content in SW480 and SW620 cell lines

The Supplementary Figure 6 shows the *TET* knockout strategy devised together with Horizon Discovery to generate single, double and triple *TET* knockouts using the CRISPR/Cas9 system. The internal validation of the data provided by Horizon Discovery was conducted by performing sequencing of all the knockout cell lines and confirming the presence of required deletions (Supplementary Figure 7). Based on the literature search on cell survival following *TET* deletions, it was decided to sequentially target *TET2*, then *TET3* and finally *TET1* in both cell lines. Since the insertions and deletions made by the CRISPR/Cas9 method can affect the stability of mRNA and lead to nonsense mediated decay (Tuladhar et al., 2019), the mRNA levels of the three TETs in each of the congenic cell lines were assessed by qRT-PCR (Figure 5.1). In the SW480 cells, the levels of *TET1*, *TET2* and *TET3* were significantly reduced in all the knockout cell lines, including the SKO and DKO cells (Figure 5.1 A). This unexpected result showing reduced *TET3* and *TET1* levels in a single *TET2*-targeted cell line could possibly be explained by the TET enzymes regulating each other's expression in the SW480 cells. The levels of *TET1* were reduced to between 0.34 and 0.41 relative to the abundance in the wild type cell line, with the values for *TET2* ranging from 0.15 to 0.20 and for *TET3* between 0.29 and 0.32. *TET2* was the most abundantly expressed transcript of the three TETs in this cell line prior to the knockout (section 4.4.1, Figure 4.2 B). The reduction observed in the knockout cell lines, while substantial, resulted in a large amount of the *TET2* transcript persisting at a level of approximately equivalent to 0.3 picograms of the standard curve vector used. This amount of transcript is approximately twice as high as the original absolute amount of *TET1* mRNA in this cell line. The total absolute TET levels were reduced to amount equivalent to approximately 0.6 picograms of vector used in the two TKO cell lines compared to the original value of 2.5 picograms in the wild type cells.

In the SW620 cell line, the mRNA level of *TET2* was no different in the SKO cells and appeared to be increased following the double knockout, when *TET3* was subsequently targeted. Given the low initial abundance of *TET2* mRNA in this cell line (section 4.4.1,

Figure 4.2 B), it appears that the CRISPR/Cas9-induced mutation did not affect the stability of the *TET2* transcript in any of the three knockout cell lines. The upregulation of the *TET2* mRNA levels in the DKO cells is in line with the unpublished observations made by the Murrell group that *TET2* is upregulated by targeting *TET3* with siRNA in HCT116 cells (data not shown). This may suggest that the TET enzymes may transcriptionally regulate one another. However, no published reports of similar results could be found and the changes in *TET2* abundance may simply be an effect of low transcription levels and background noise. In the TKO cells, the *TET1* transcript were significantly reduced (lowered to 0.39 and 0.30, respectively, of the SKO and DKO levels) (Figure 5.1 B). Based on the absolute *TET* levels in the SW620 cell line (section 4.4.1, Figure 4.2 B), the total residual *TET* transcript levels in the TKO cell line are calculated at equivalent to approximately 0.7 picograms of vector used compared to 1.4 picograms in the wild type cells.

The levels of the three TET proteins in the knockout SW480 and SW620 cell lines were assessed by Western blotting (Figure 5.2). The representative image of the antibody staining is shown in Figure 5.2 A. In the SW480 knockout cell lines, TET1 content was found to be reduced by 78% in TKO 1 cells compared to the wild type cells (Figure 5.2 B). The reduction in TET3 content was the highest out of all the TETs, reaching 86% in the TKO 1 cell line and 92% in the TKO 2 cells compared to the abundance observed in the SW480 wild type line.

The TET protein levels in the three SW620 knockout cell lines were not reduced as efficiently as in the primary cell line (Figure 5.2 C). None of the changes in the protein levels of the three TET enzymes were found to be significant. It should be noted that the variation in TET3 signal obtained for the three independent samples analysed was higher than the other TETs. It is hence likely that this counterintuitive result is due to experimental error or variation in antibody staining as the anti-TET3 antibody produced a number of non-specific bands (see section 4.4.1, Figure 4.3).

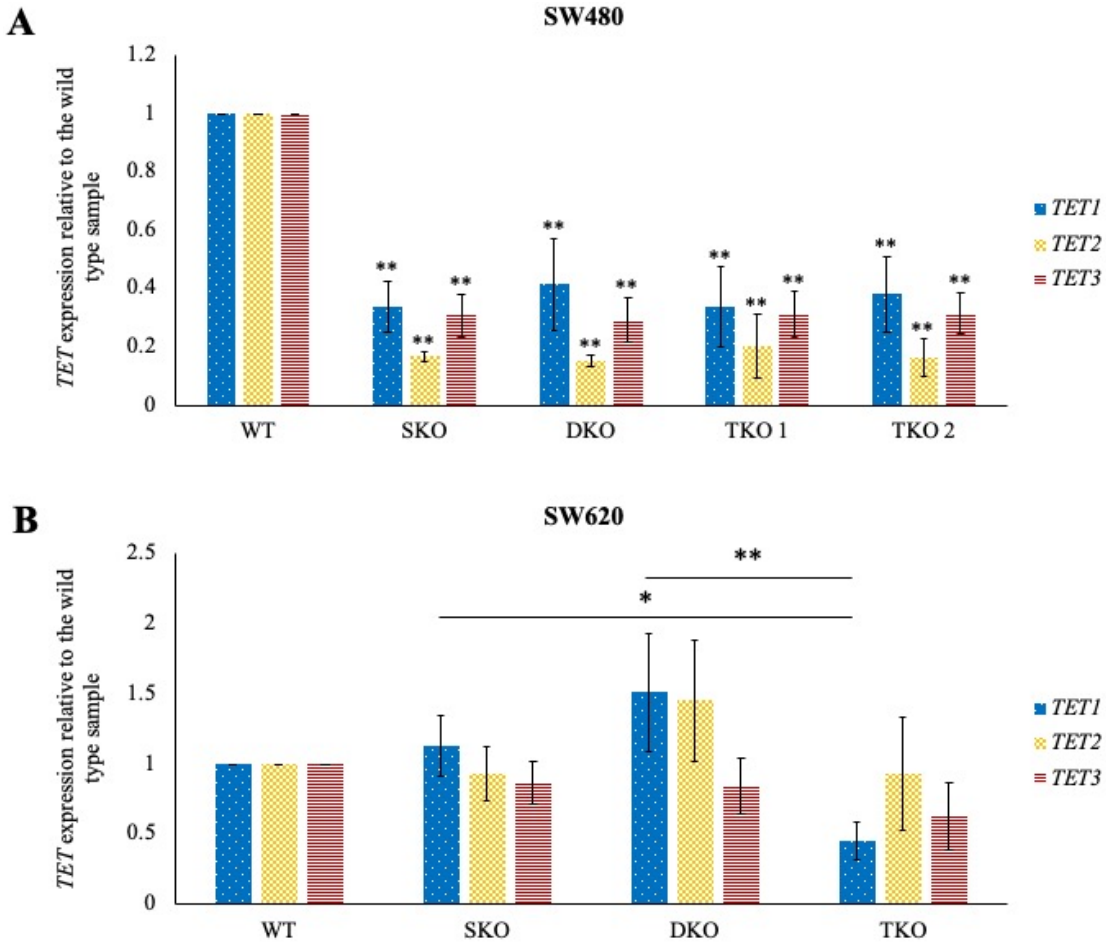


Figure 5.1 The CRISPR/Cas9 genetic knockout strategy failed to ablate the expression of the *TET* genes in the SW480 and SW620 cell lines

The expression of the *TET* transcripts in (A) SW480 and (B) SW620 wild type (WT) and single (SKO), double (DKO) and triple (TKO) *TET* knockout cell lines. The *TET* mRNA levels were normalised to *ACTB* reference gene. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$; ** $p<0.01$).

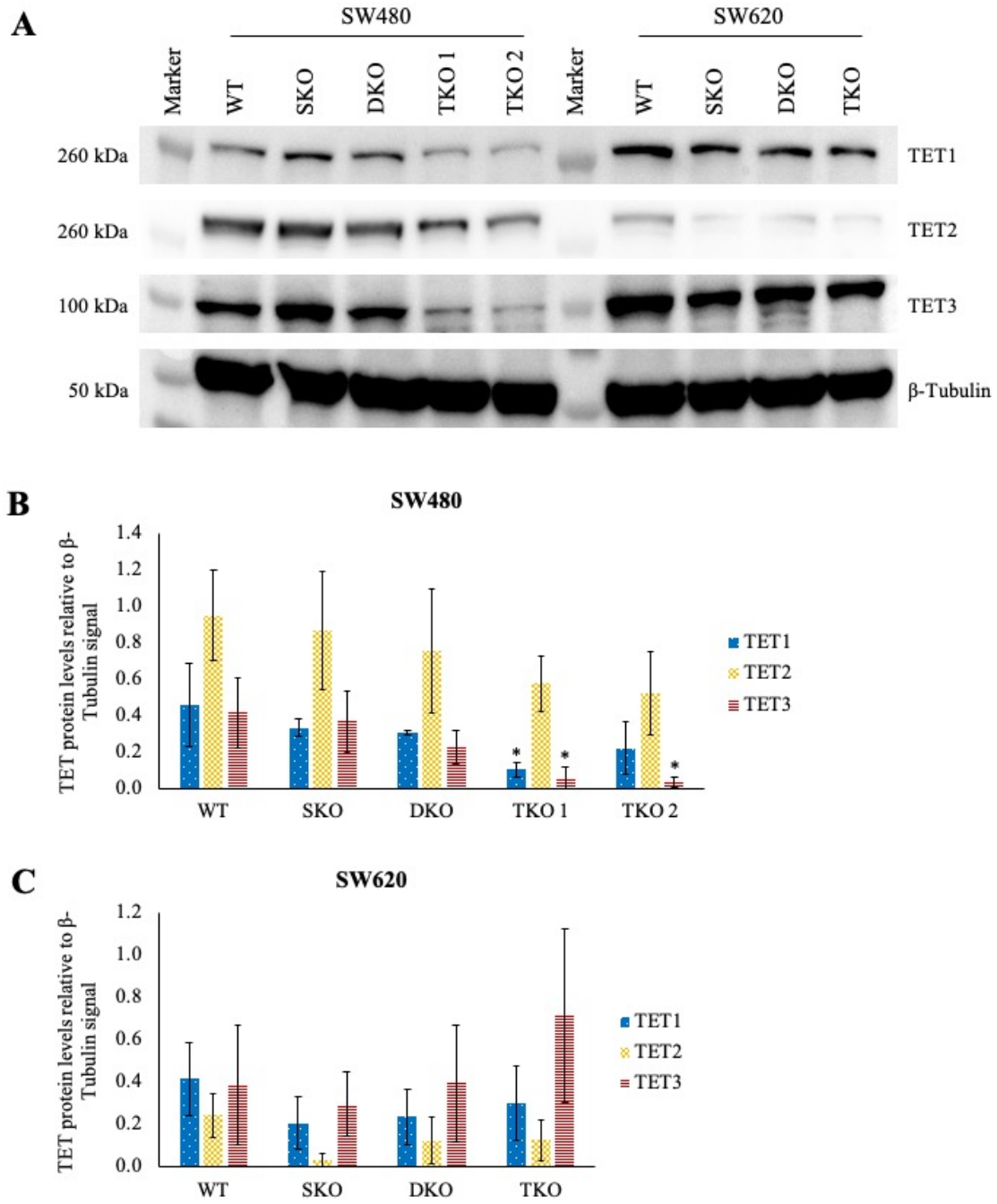


Figure 5.2 The CRISPR/Cas9 genetic knockout strategy failed to ablate the TET protein levels in the SW480 and SW620 cell lines

(A) A representative Western blot image of the assessment of TET1, TET2, TET3 and ß-Tubulin protein levels in wild type (WT), single (SKO), double (DKO) and triple (TKO) *TET* knockout SW480 and SW620 cell lines.

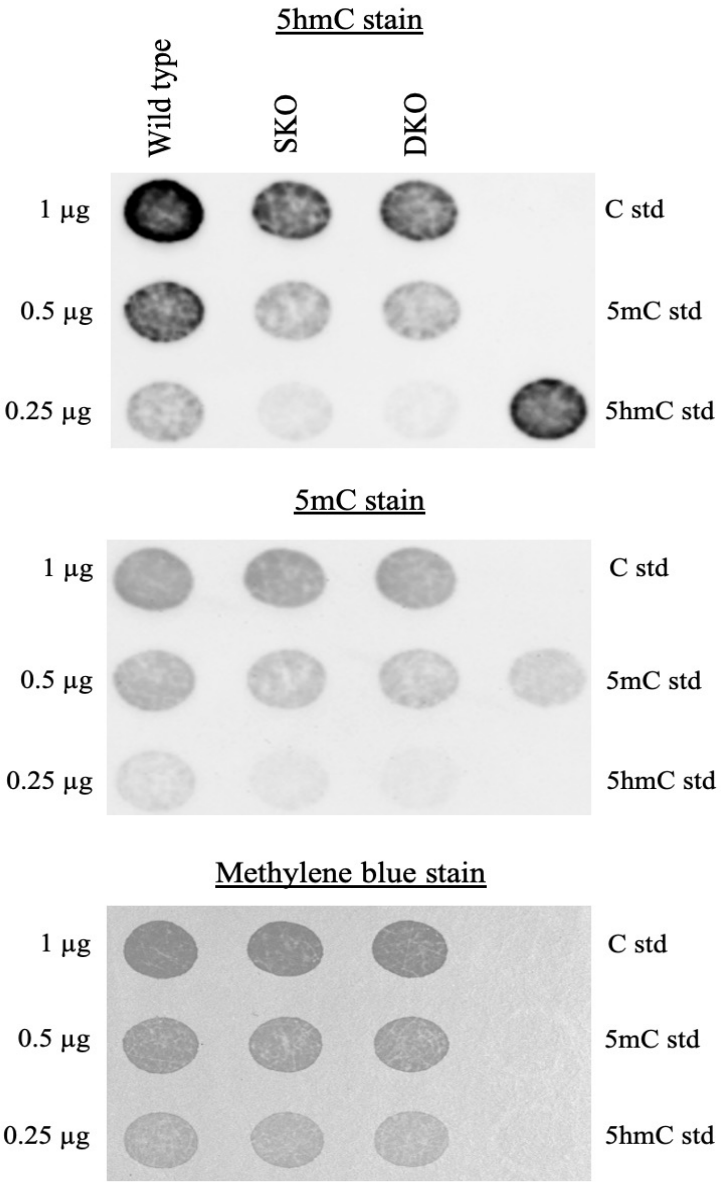
The TET protein levels in WT, SKO, DKO and TKO (B) SW480 and (C) SW620 cell lines. The staining results were normalised to the ß-Tubulin content in each cell line. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$).

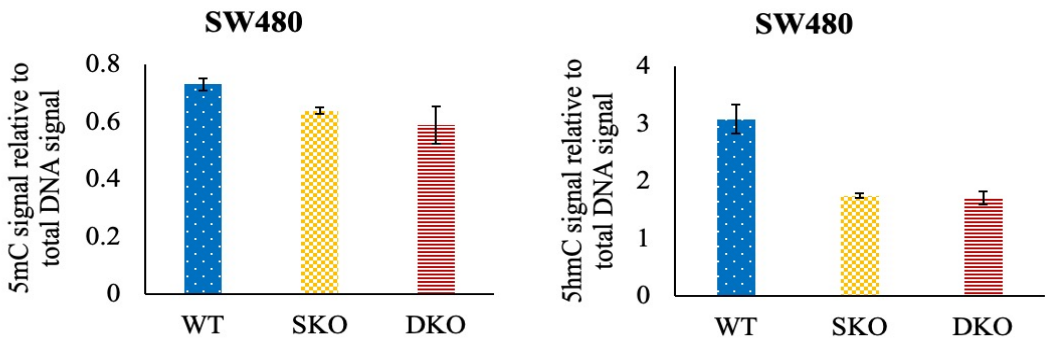
The *TET* knockout SW480 and SW620 cell lines were also assessed for their global 5mC and 5hmC content using the dot blot technique (Figure 5.3). The single and double knockout SW480 cell lines both have a modest reduction in 5mC levels and a similar approximately 40% reduction in 5hmC content compared to the wild type cells (Figure 5.3 A). The single and double knockout SW620 cells on the other hand display no changes in 5mC levels with global 5hmC levels lower by approximately 15% relative to the unmodified SW620 cell line (Figure 5.3 B). The triple *TET* knockout in SW480 cells resulted in no change in 5mC levels and a 65% reduction in 5hmC content for both knockout cell lines (Figure 5.3 C). The triple TET knockout SW620 cell line showed slightly higher levels of 5mC and approximately 25% lower 5hmC levels compared to the wild type SW620 cells (Figure 5.3 D). Taken together, this data points to a more efficient knockdown of TET transcript and protein levels, as well as the activity of these enzymes, in the primary SW480 cell line compared to its metastatic SW620 counterpart. This is further supported by the TET activity assay results shown in Supplementary Figure 8.

The mass spectrometry data for the triple *TET* knockout cell lines is not shown due to the 5hmC signal having a signal to noise ratio below 10 and hence being non-quantifiable (data not shown). The levels of 5mC were found to be significantly increased in the TKO SW620 cells, but not in TKO1 and TKO 2 SW480 cell lines, compared to the wild type cells as measured by mass spectrometry (Supplementary Figure 9). This could be due to the aforementioned reduced global 5hmC content in this cell line.

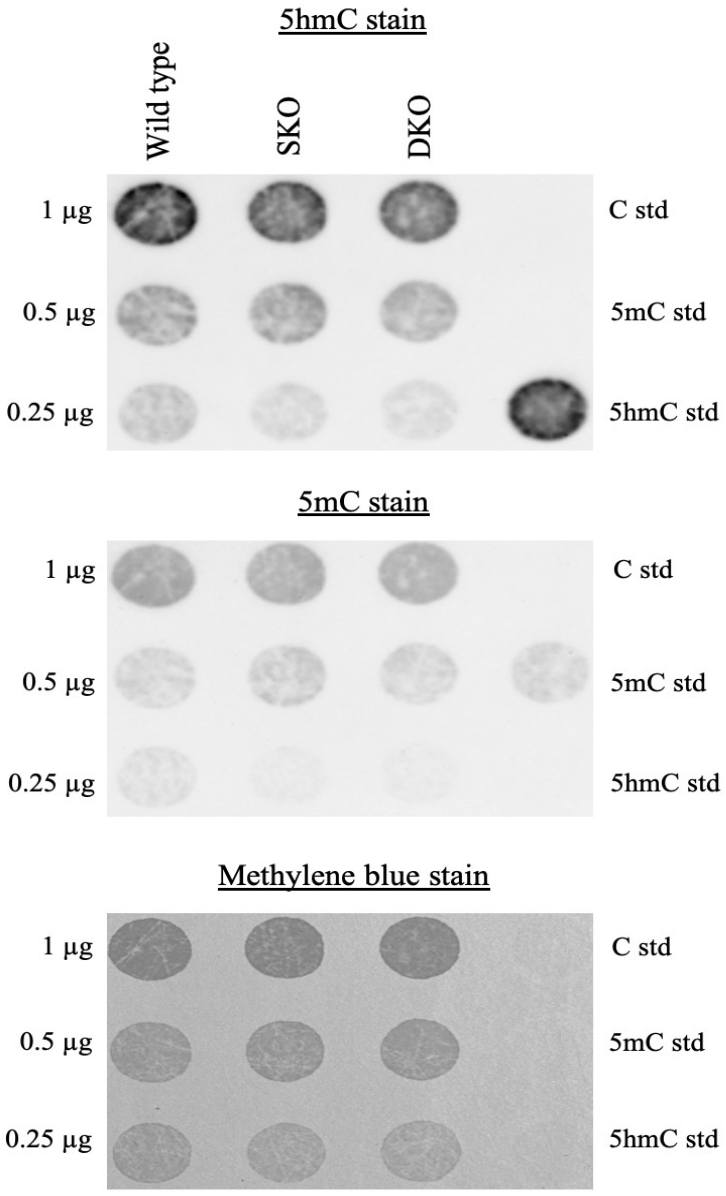
Based on the results shown in Figure 5.1, Figure 5.2 and Figure 5.3 the cell lines will be referred to in subsequent text as knockdowns (KD) rather than knockouts (KO) to accurately describe their identity.

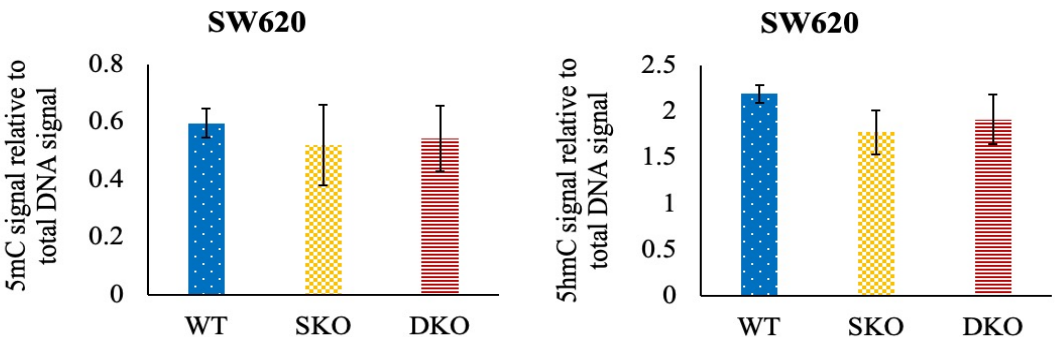
A


5hmC stain
Wild type
SKO
DKO
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
5mC stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
Methylene blue stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std

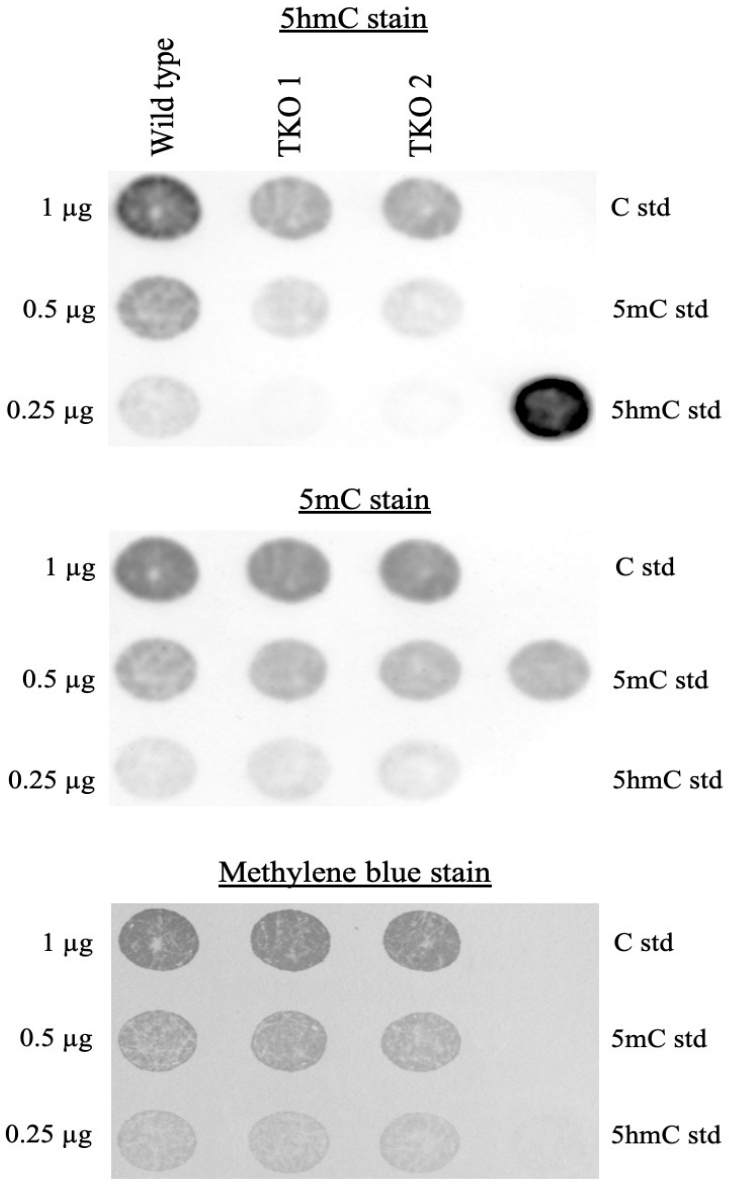

SW480
5mC signal relative to total DNA signal
0.8
0.6
0.4
0.2
0
WT
SKO
DKO
SW480
5hmC signal relative to total DNA signal
4
3
2
1
0
WT
SKO
DKO

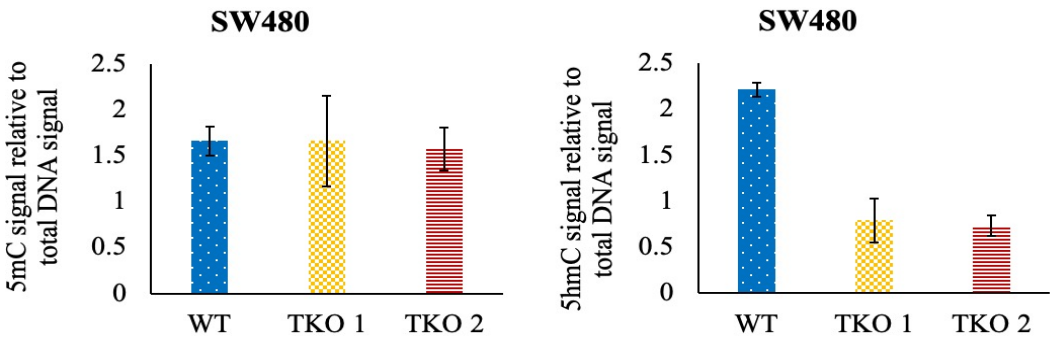
B


5hmC stain
Wild type
SKO
DKO
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
5mC stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
Methylene blue stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std


SW620
5mC signal relative to total DNA signal
0.8
0.6
0.4
0.2
0
WT
SKO
DKO
SW620
5hmC signal relative to total DNA signal
2.5
2
1.5
1
0.5
0
WT
SKO
DKO

C


5hmC stain
Wild type
TKO 1
TKO 2
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
5mC stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
Methylene blue stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std


SW480
5mC signal relative to total DNA signal
2.5
2
1.5
1
0.5
0
WT
TKO 1
TKO 2
SW480
5hmC signal relative to total DNA signal
2.5
2
1.5
1
0.5
0
WT
TKO 1
TKO 2

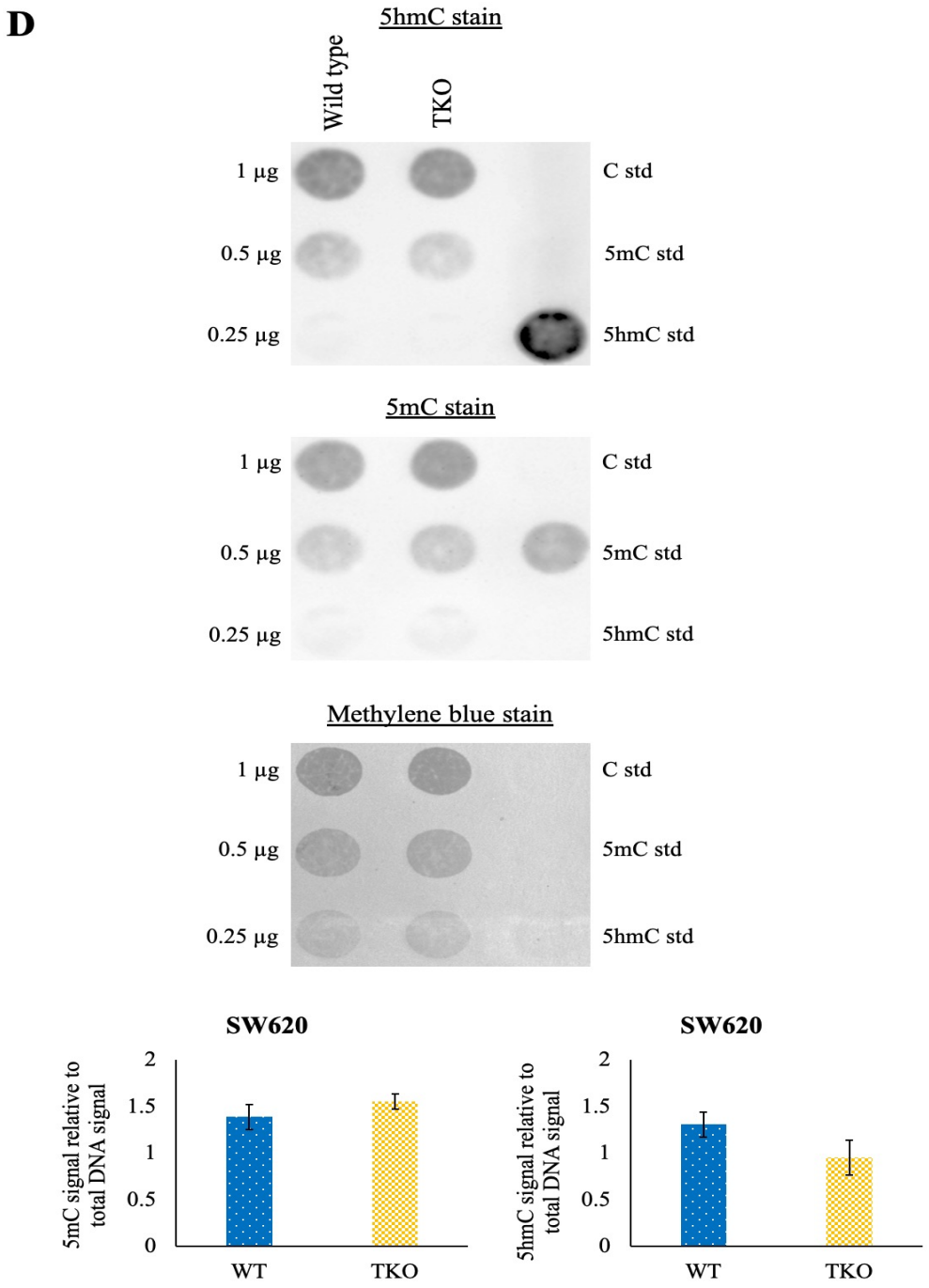


Figure 5.3 The CRISPR/Cas9 genetic knockout strategy failed to ablate the global 5hmC content in the SW480 and SW620 cell lines

(A) and (B) The representative staining and graphical representation of the results of dot blot assessment of 5hmC and 5mC levels in wild type (WT), single (SKO) and double (DKO) *TET* knockout SW480 and SW620 cell lines, respectively. The errors bars represent the standard deviation of three independent replicates.

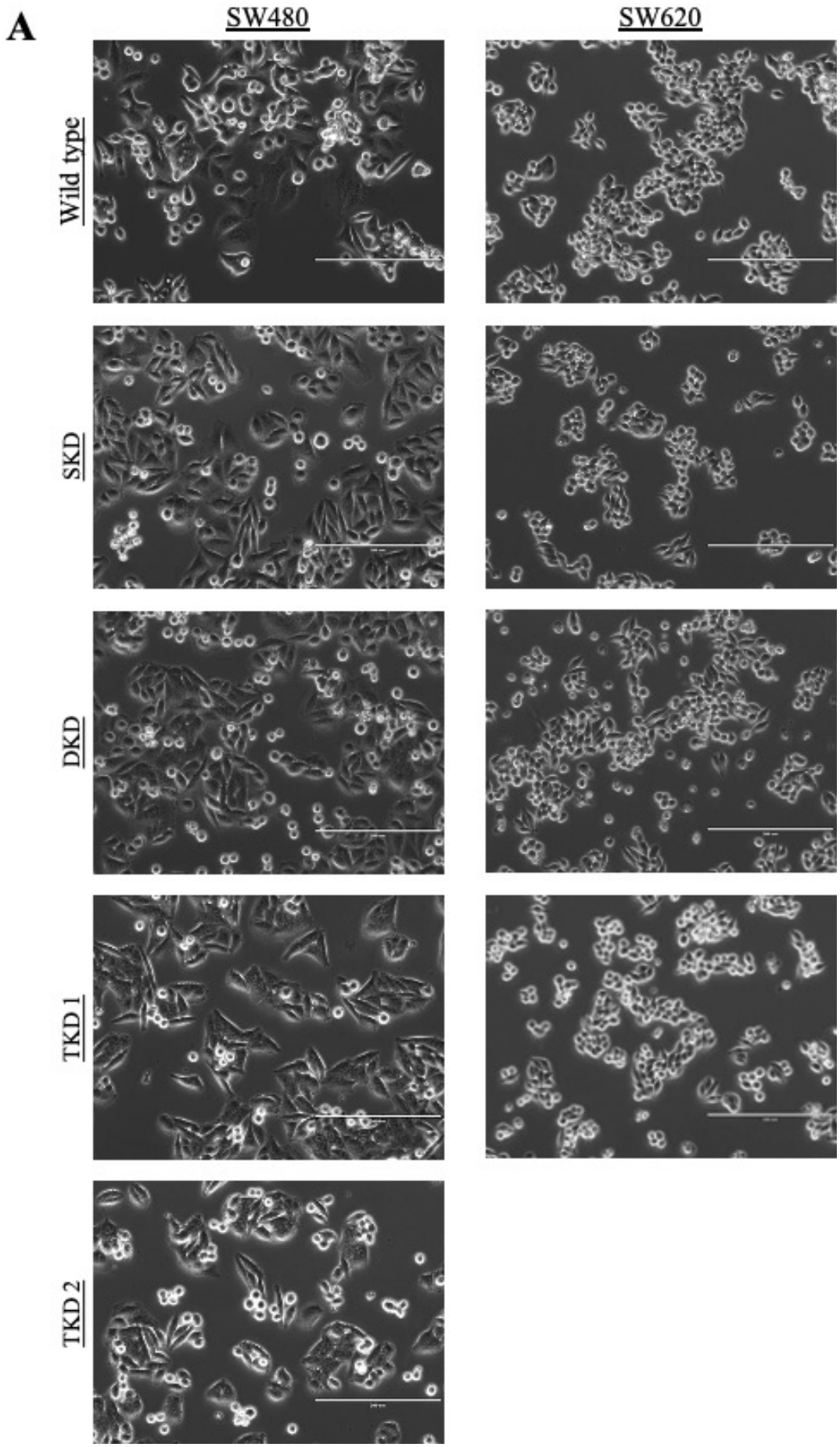
(C) and (D) The representative staining and graphical representation of the results of dot blot assessment of 5hmC and 5mC levels in WT and triple (TKO) *TET* knockout SW480 and SW620 cell lines, respectively. The errors bars represent the standard deviation of four independent replicates.

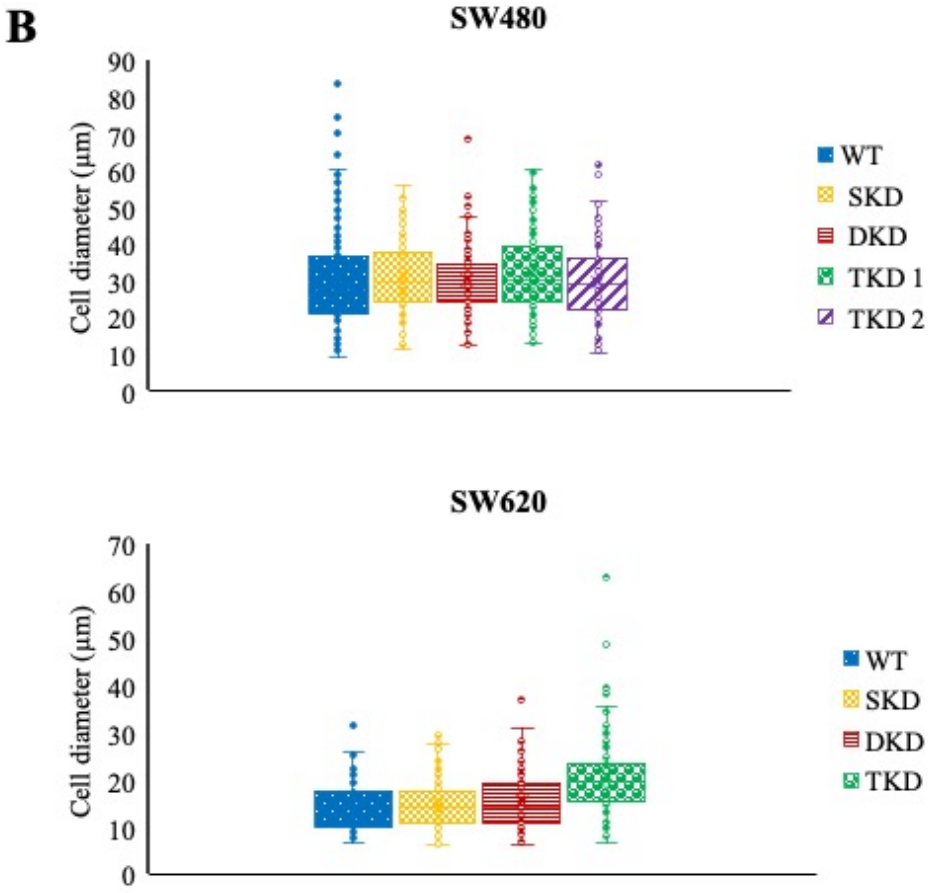
5.4.2 The impact of the *TET* knockdowns on cell proliferation, colony formation and migration in the SW480 and SW620 cells

Based on the results described above, the TKD SW480 and SW620 cell lines are not a complete functional knockout of the TET enzymes. Nevertheless, the reduction in the TET levels and their catalytic activity puts these cell lines on par with multiple knockdowns of the TET enzymes described in the literature (Wu et al., 2015; Cui et al., 2016; Pei et al., 2016; Zhou et al., 2018). Based on the previously reported phenotypic changes due to the said knockdowns, it is likely that there will be pronounced phenotypic differences between the TKOD and wild type SW480 and SW620 cell lines.

The SW480 and SW620 cell lines were next assessed for the effect of the TET enzymes knockdown on cell morphology and size (Figure 5.4). The SKD, DKD and TKD cell lines did not exhibit any obvious morphological differences compared to the wild type cells (Figure 5.4 A). The SW480 knockdown cell lines were also found to have no difference in cell size compared to the wild type cells (Figure 5.4 B). On the other hand, the SKD and DKD SW620 cell lines have significantly reduced cell diameter, 15.1 and 15.6 μm, compared to the wild type cells (17.3 μm) while the TKD SW620 cells were found to have a significantly higher cell diameter of 21 μm (Figure 5.4 B).

The rate of cell proliferation was compared between the wild type and knockdown SW480 and SW620 cell lines (Figure 5.5). Only one of the *TET* knockdown SW480 cell lines (TKD 2) displayed a significantly reduced proliferation rate compared to the wild type cells at 72-hour (615,000 and 900,000 cells, respectively) and compared to the TKD1 cell line at 96-hour (933,000 and 1,294,000 cells, respectively) time points (Figure 5.5 A). None of the *TET* knockdowns resulted in a significantly altered rate of cell division in the SW620 cell line (Figure 5.5 B).


A
SW480
SW620
Wild type
SKD
DKD
TKD 1
TKD 2



Sample comparison (cell diameter)	Significance level
WT vs SKD	*
WT vs DKD	ns
WT vs TKD	**
SKD vs DKD	ns
SKD vs TKD	**
DKD vs TKD	**

Figure 5.4 The cell size comparison between wild type and *TET* knockdown SW480 and SW620 cells

(A) The representative images of SW480 and SW620 wild type (WT), single (SKD), double (DKD) and triple (TKD) *TET* knockdown cell lines taken 48 hours after seeding at 20X magnification. Scale bar = 200 µm.

(B) The comparison of cell diameter between the wild type and *TET* knockdown SW480 and SW620 cell lines. The error bars represent the standard deviation of two hundred wild type one hundred knockdown cells measured for each cell line in a single culture dish. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$; ** $p<0.01$). The results of the statistical test are shown in the table below the graph for the SW620 cells.

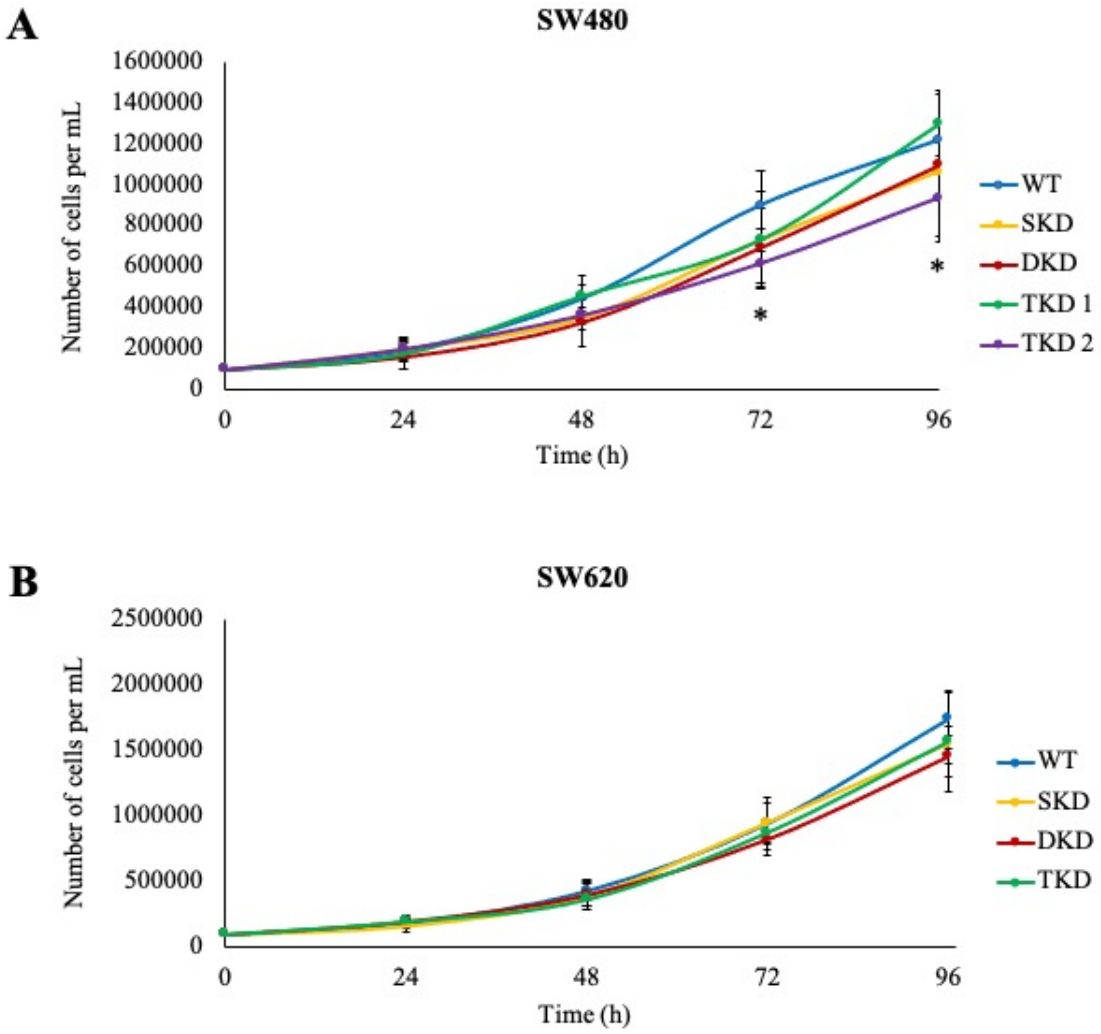


Figure 5.5 The comparison of proliferation rates of wild type and *TET* knockdown SW480 and SW620 cell lines

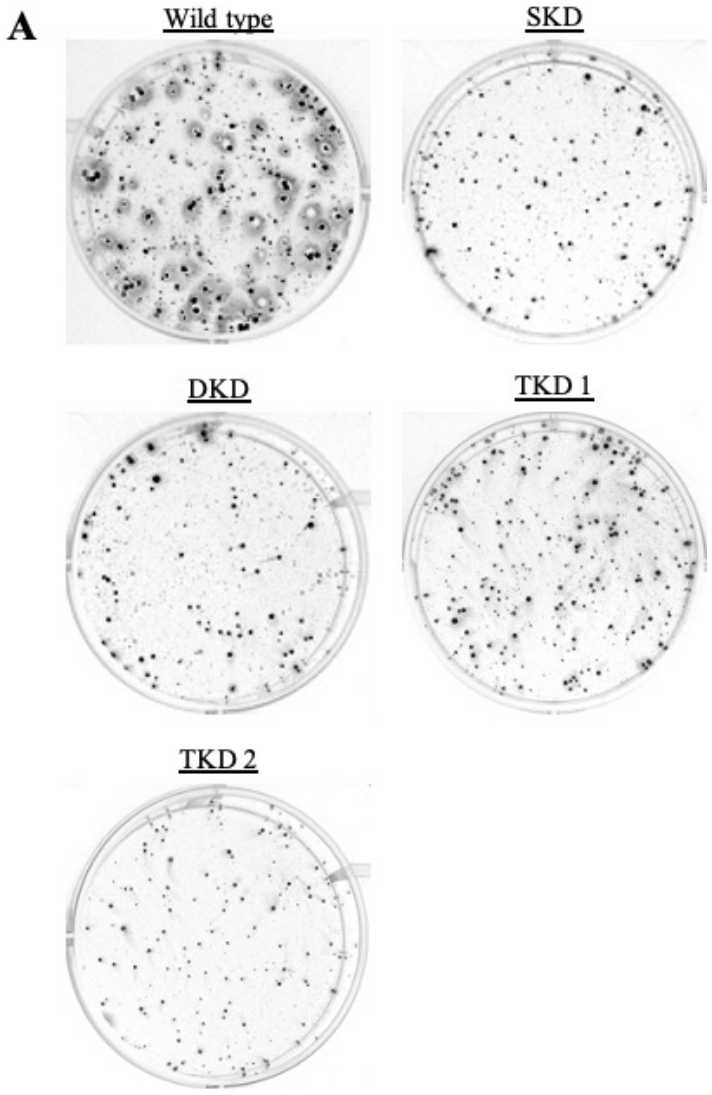
The cell counts of (A) SW480 and (B) SW620 wild type (WT), single (SKD), double (DKD) and triple (TKD) *TET* knockdown cell lines measured at 24 hour intervals for 96 hours. The error bars represent the standard deviation of eight independent replicates for the WT cells, three independent replicates for the SKD and DKD cells and five independent replicates for the TKD cells. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$).

The colony formation assay with the abovementioned cell lines provided a different insight into their rate of *in vitro* growth (Figure 5.6). The colony formation appeared to be reduced for almost all TET knockdowns in both SW480 (Figure 5.6 A) and SW620 (Figure 5.6 B) cell lines. The number of colonies formed was significantly reduced in both the SKD and DKD SW480 cell lines (reduction in the average colony count of 56 and 53%, respectively) compared to the wild type cells (Figure 5.6 C). Additionally,

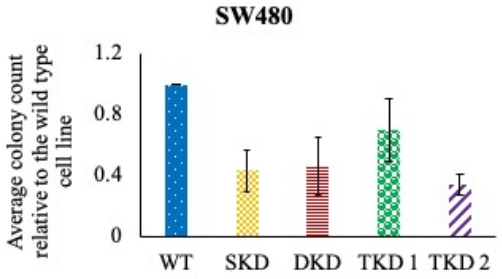
SKD SW480 cells showed a significantly smaller average colony area (reduced by 30%) than the unmodified cells. TKD 2 SW480 cells displayed a reduced colony formation ability when compared to the wild type cell line (reduction in the average colony count of 65). The TKD 1 SW480 cell line formed 25% larger colonies than the wild type cells with no significant changes observed in the TKD 2 cells.

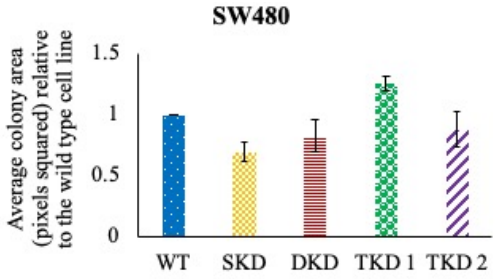
The SKD, DKD and TKD SW620 cells were all found to have a significantly reduced number of the colonies formed compared to the unmodified cell line (by 32, 42 and 47%, respectively) (Figure 5.6 D). The DKD and TKD SW620 cell lines were also found to have a significantly reduced colony size (by 12 and 23%, respectively) compared the wild type colonies.

The migration of the wild type and *TET* knockdown SW480 and SW620 was assessed by the scratch assay (Figure 5.7). The migration rate appeared higher in the knockout SW480 cell lines compared to unmodified cells (Figure 5.7 A). The opposite pattern was observed in the SW620 cell line, with the knockdown cells migrating less than the wild type cells (Figure 5.7 B). All the SW480 *TET* knockdown cell lines were found have migrated the same distance as the wild type cells inside the scratched gap with no significant changes observed (Figure 5.7 C). The SW620 *TET* knockdown cell lines, on the other hand, showed reduced migration ability compared to the parent cell line (Figure 5.7 D). The percentage gap closure was found to be significantly lower for the TKD SW620 cell line (12%) compared to the wild type cells (23%) (Figure 5.7 D).


A
Wild type
SKD
DKD
TKD 1
TKD 2
B
Wild type
SKD
DKD
TKD

C

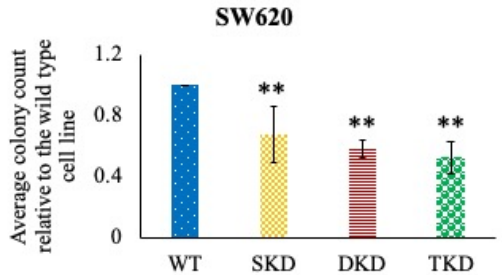




Sample comparison (colony count)	Significance level
WT vs SKD	**
WT vs DKD	**
WT vs TKD 1	ns
WT vs TKD 2	**
SKD vs DKD	ns
SKD vs TKD 1	ns
SKD vs TKD 2	ns
DKD vs TKD 1	ns
DKD vs TKD 2	ns
TKD 1 vs TKD 2	*

Sample comparison (colony area)	Significance level
WT vs SKD	**
WT vs DKD	ns
WT vs TKD 1	*
WT vs TKD 2	ns
SKD vs DKD	ns
SKD vs TKD 1	**
SKD vs TKD 2	ns
DKD vs TKD 1	**
DKD vs TKD 2	ns
TKD 1 vs TKD 2	**

D



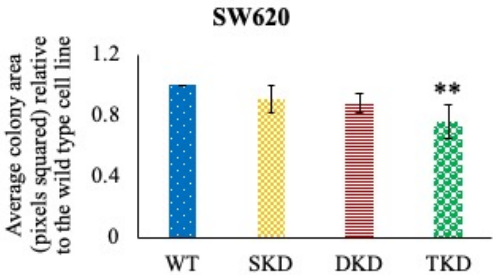
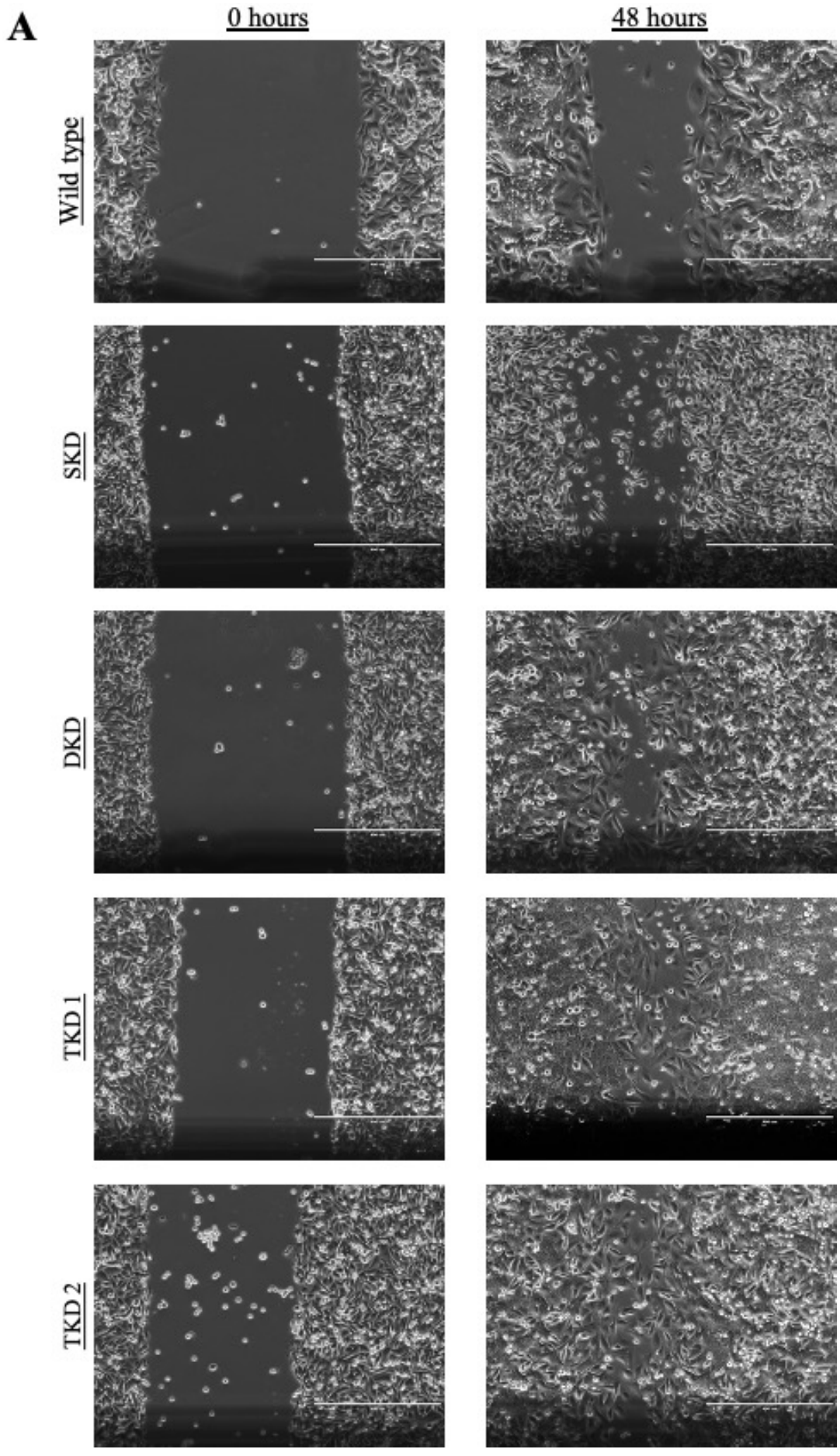
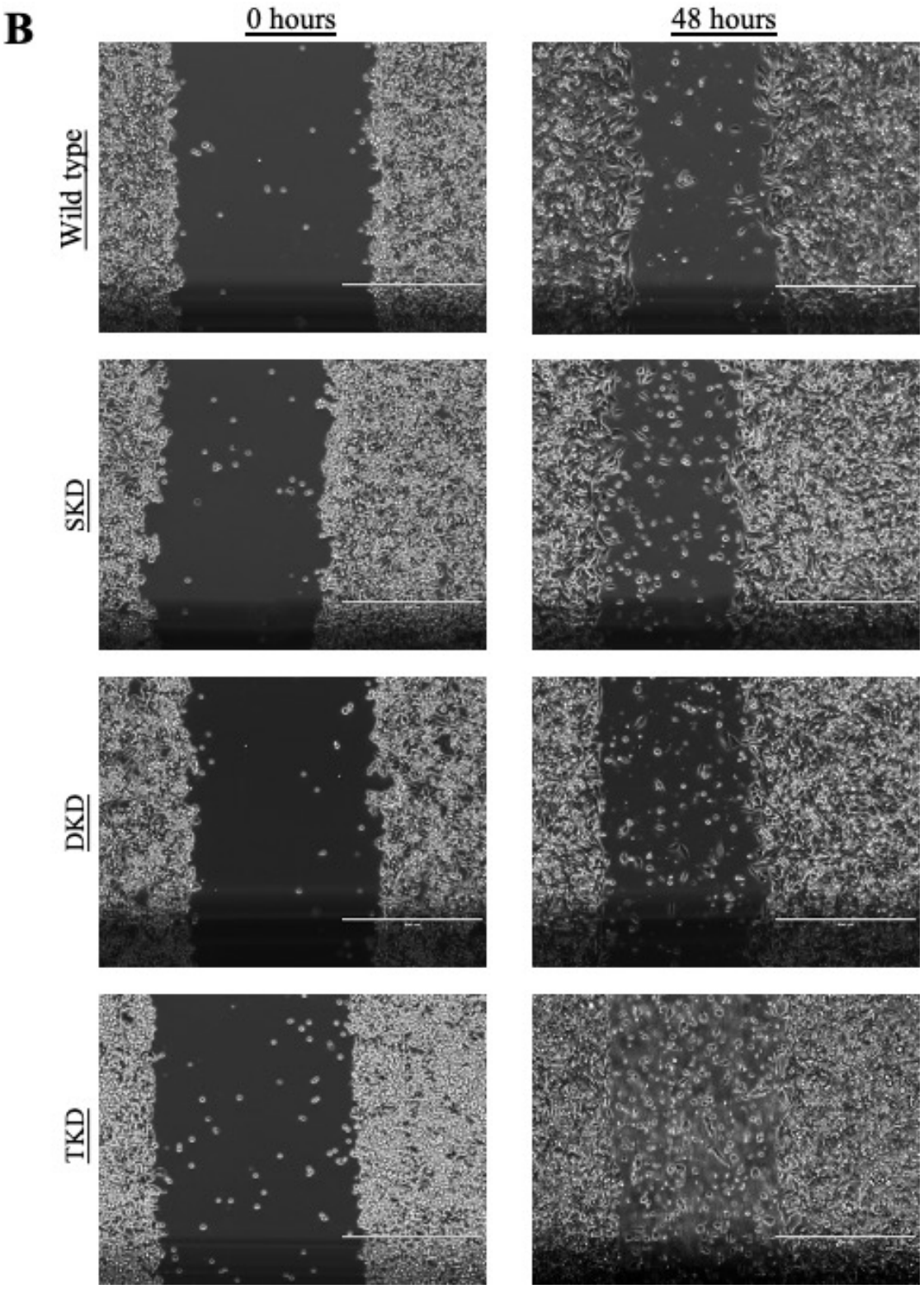


Figure 5.6 The knockdowns of the TET enzymes lead to a reduction in colony formation ability of SW480 and SW620 cells

(A) and (B) The representative images of colony formation of wild type (WT), single (SKD), double (DKD) and triple (TKD) *TET* knockdown SW480 and SW620 cell lines, respectively, taken between 10 and 14 days after seeding.

(C) and (D) The depiction of the average colony size and area in WT, SKD, DKD and TKD SW480 and SW620 cell lines, respectively. The error bars represent the standard deviation of four independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$; ** $p<0.01$). (C) The results of the statistical test are shown in the table below the graph.


A
0 hours
48 hours
Wild type
SKD
DKD
TKD 1
TKD 2


B
0 hours
48 hours
Wild type
SKD
DKD
TKD

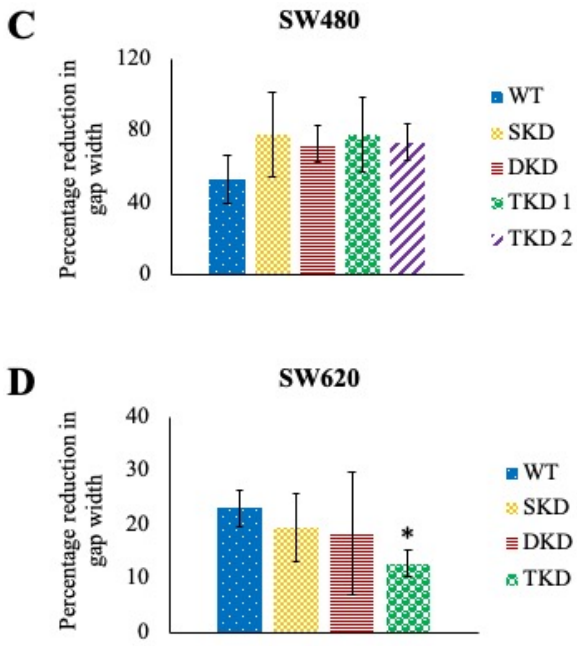


Figure 5.7 The triple knockdown of the TET enzymes lead to a reduction in migration of the SW620 cell line

(A) and (B) The representative images of migration of wild type (WT), single (SKD), double (DKD) and triple (TKD) *TET* knockdown SW480 and SW620 cell lines, respectively. The images were taken at 0 hours and 48 hours after the scratch was made, at 10X magnification. Scale bar = 400 µm.

(C) and (D) The depiction of the percentage gap closure between the 0 hour and 48 hour time points in WT, SKD, DKD and TKD SW480 and SW620 cell lines, respectively. The error bars represent the standard deviation of eight independent replicates for the WT cells, three independent replicates for the SKD and DKD cells and five independent replicates for the TKD cells. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$).

5.4.3 Assessment of the effect of the TET enzymes levels rescue through acute overexpression on the migration of the triple knockdown SW480 and SW620 cell lines

The results showing pronounced effects of *TET* triple knockdown on migration of SW620 cells were followed up by a rescue experiment using transient overexpression of each of the TET enzymes in the three TKD cell lines. While the changes in migration observed with *TET* knockdown SW480 cell lines were found to not be significant, they all appeared to have an elevated migration rate compared to the wild type cells. The TKD 1 and TKD 2 cell lines were hence also included in this investigation. This experiment would allow one to understand better which of the TET enzymes had the biggest effect on migration of the two CRC cell lines. The overexpression constructs used contain human *TET1* and *TET3* genes and mouse *Tet2* gene (Tahiliani et al., 2009; Ko et al., 2010; Ko et al., 2013).

The overexpression of the three TET enzymes in all the transfected samples was validated by qRT-PCR (Figure 5.8). The expression of *TET1* was on average 19-fold higher compared in the TET1 overexpression sample to the control TKD 1 SW480 cells (Figure 5.8 A). The levels of *TET2* did not change in any of the samples, indicating no positive or negative feedback on its expression due to the acute overexpression of the other TETs. The successful expression of the mouse *Tet2* transcript was confirmed in the transfected cells (see Supplementary Figure 10 for *Tet2* expression relative to the *ACTB* levels), and *TET3* levels were found to be approximately10-fold higher in the TET3 overexpression sample compared to controls.

In the TKD 2 SW480 cell line, the *TET1* levels were on average 23-fold higher in the TET1 overexpressing cells compared to the control samples (Figure 5.8 B). In a similar fashion to the results obtained in the TKD 1 SW480 cells (Figure 5.8A), the *TET2* expression remained largely unchanged in all samples. The successful expression of mouse *Tet2* transcript was confirmed in the transfected cells (see Supplementary Figure 10 for *Tet2* expression relative to the *ACTB* levels), and *TET3* levels were over 7-fold higher in the TET3 overexpression sample compared to controls.

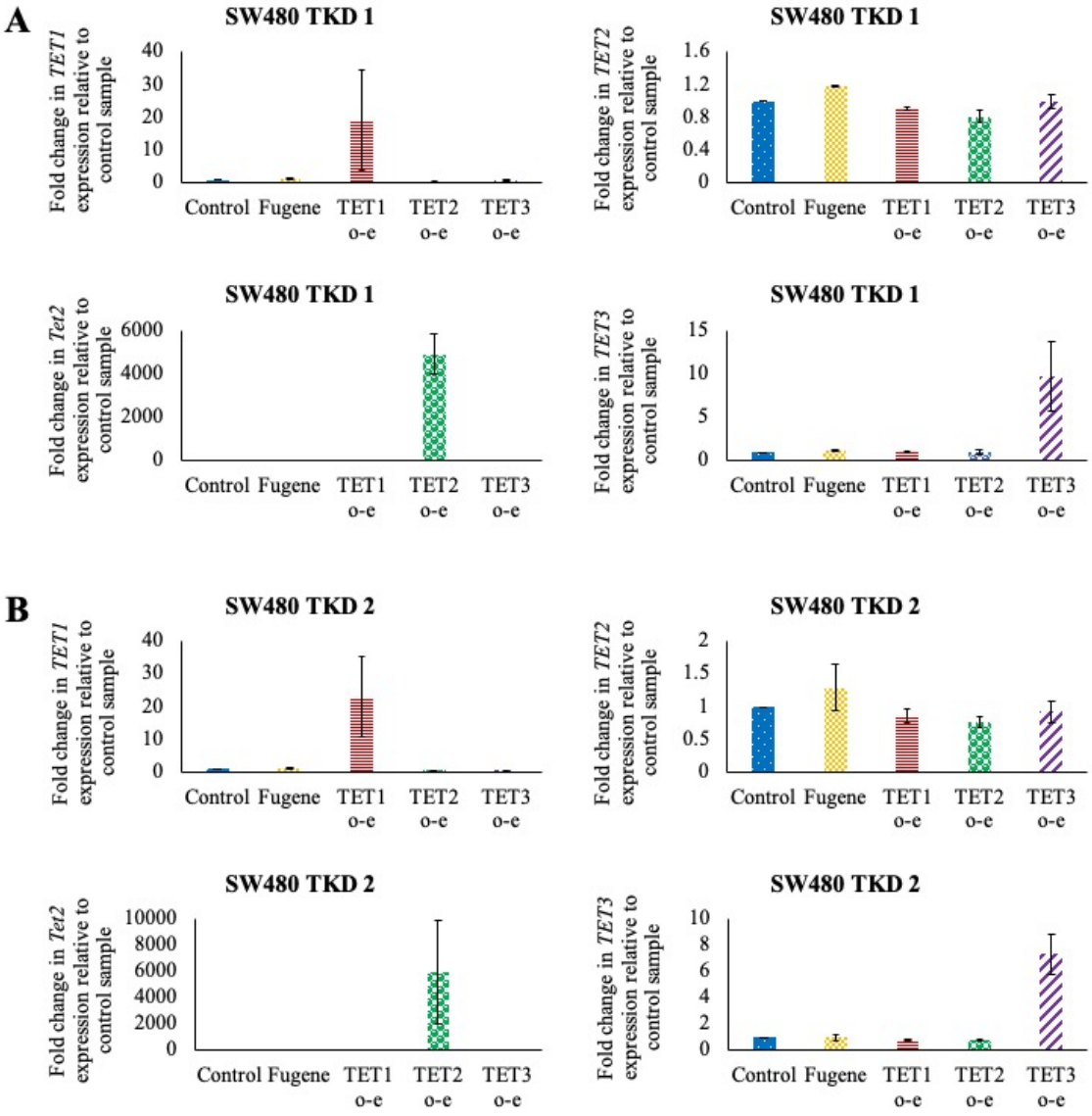
The *TET1* levels were on average over 6-fold higher in the TKD SW620 TET1 overexpressing cells compared to the control samples (Figure 5.8 C). The levels of *TET2* were unchanged by the overexpression of the three TET enzymes and the

successful expression of mouse *Tet2* transcript was also confirmed (see Supplementary Figure 10 for *Tet2* expression relative to the *ACTB* levels). Lastly, *TET3* levels were found to be approximately 6-fold higher in the TET3 overexpressing cells compared to the untransfected control.

The overexpression of each of the three TET enzymes led to a modest decrease in the migration of the TKD 1 SW480 cell line with no obvious increase in the cell death due to the transfection process (Figure 5.9 A). Overexpression of the TET3 enzyme was found to have the greatest effect on migration and resulted in significantly reduced gap closure compared to all the other samples (Figure 5.9 B). The overexpression of TET1 and TET3 resulted in 22 and 48% reduction in gap closure compared to the untransfected cells. Additionally, TET3 overexpression led to a 44% lower reduction in gap width compared to cells treated with transfection reagent-only.

The individual TET overexpression in TKD 2 SW480 cells, in a similar fashion to the TKD 1 cell line, resulted in a small reduction in their migration ability compared to the control cell lines (Figure 5.10 A). However, in contrast to the TKD 1 cells, acute overexpression of the three TET enzymes resulted no change in the migration ability of the TKO 2 SW480 cells (Figure 5.10 B).

The TKO SW620 cells similarly appeared to have a lower migration rate when the three TET enzymes were individually overexpressed exogenously (Figure 5.11 A). All overexpression samples exhibited a reduced percentage gap closure compared to the untransfected control and TET3 overexpression also significantly lowered the percentage gap closure compared to transfection only control (Figure 5.11 B). TET3 overexpressing cells were determined to close the scratch gap the least of the three overexpression samples (4%), followed by the TET2 (9%) and TET1 (10%) overexpressing cells.


A
SW480 TKD 1
Fold change in TET1 expression relative to control sample
40
30
20
10
0
Control Fugene TET1 o-e TET2 o-e TET3 o-e
SW480 TKD 1
Fold change in TET2 expression relative to control sample
1.6
1.2
0.8
0.4
0
Control Fugene TET1 o-e TET2 o-e TET3 o-e
SW480 TKD 1
Fold change in Tet2 expression relative to control sample
6000
4000
2000
0
Control Fugene TET1 o-e TET2 o-e TET3 o-e
SW480 TKD 1
Fold change in TET3 expression relative to control sample
15
10
5
0
Control Fugene TET1 o-e TET2 o-e TET3 o-e
B
SW480 TKD 2
Fold change in TET1 expression relative to control sample
40
30
20
10
0
Control Fugene TET1 o-e TET2 o-e TET3 o-e
SW480 TKD 2
Fold change in TET2 expression relative to control sample
2
1.5
1
0.5
0
Control Fugene TET1 o-e TET2 o-e TET3 o-e
SW480 TKD 2
Fold change in Tet2 expression relative to control sample
10000
8000
6000
4000
2000
0
Control Fugene TET1 o-e TET2 o-e TET3 o-e
SW480 TKD 2
Fold change in TET3 expression relative to control sample
10
8
6
4
2
0
Control Fugene TET1 o-e TET2 o-e TET3 o-e

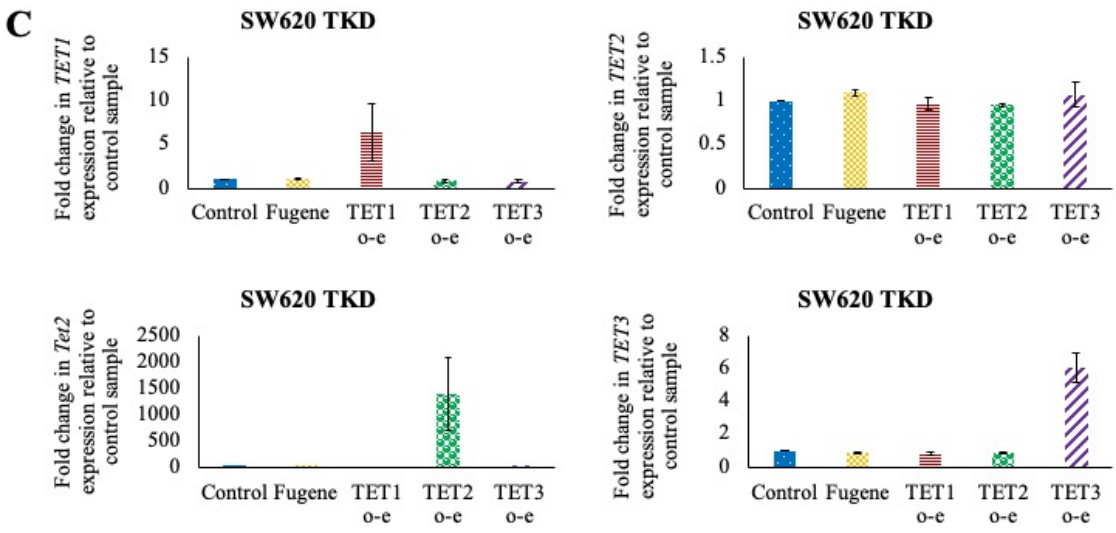
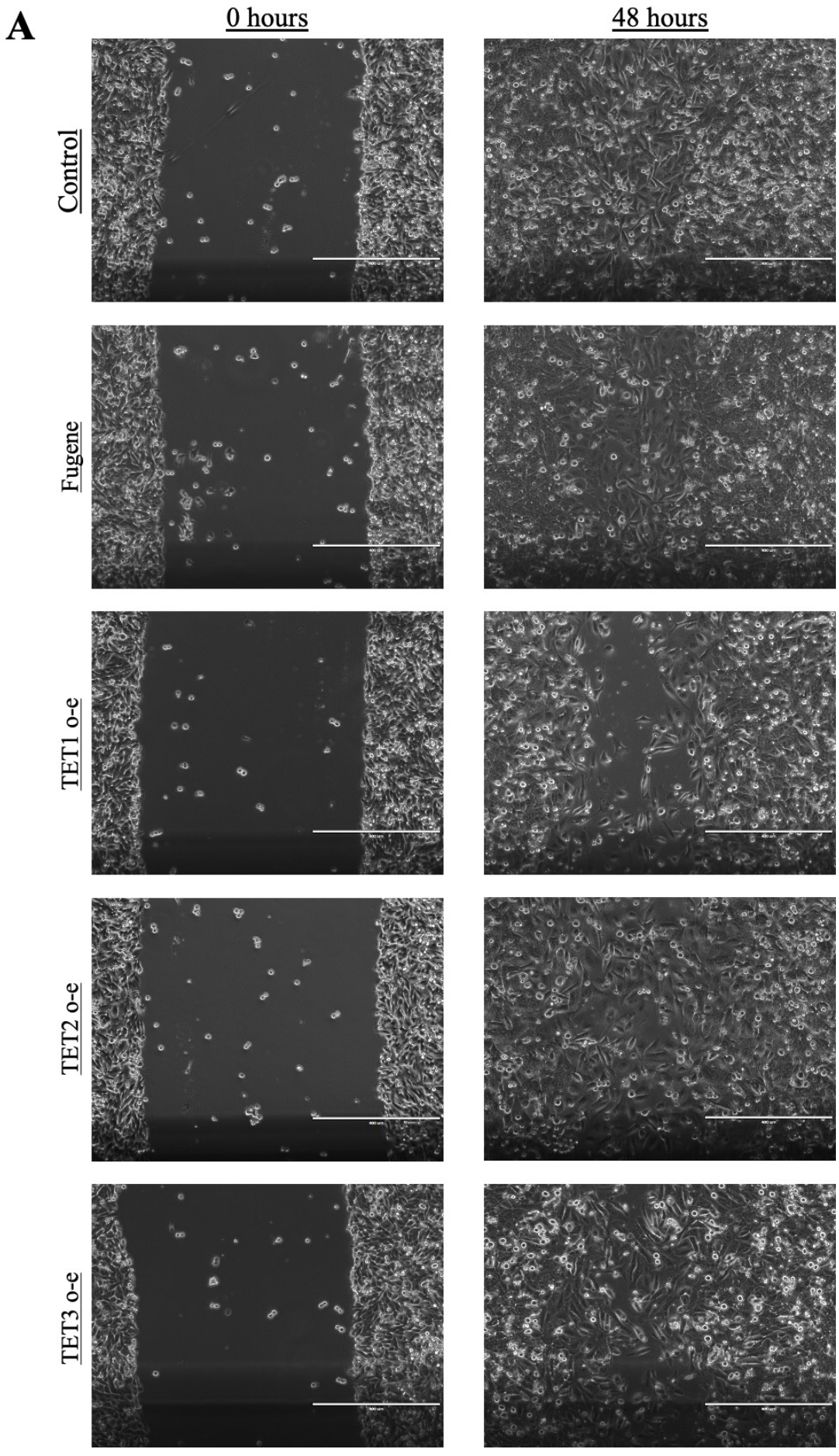
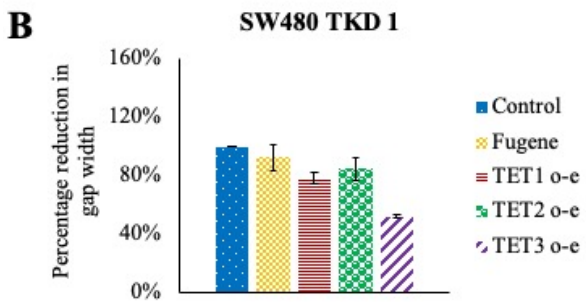


Figure 5.8 The acute TET enzymes overexpression leads to an upregulation in the corresponding mRNA levels in the three TKD cell lines
The expression of the human *TET1/2/3* and mouse *Tet2* transcripts in untransfected control, transfection reagent-only control (Fugene) and TET1/2/3 overexpressing (TET1/2/3 o-e) (A) TKD 1 SW480, (B) TKD 2 SW480 and (C) TKD SW620 cell lines. The *TET* mRNA levels were normalised to *ACTB* reference gene. The error bars represent the standard deviation of three independent replicates.

The TET3 overexpression also resulted in significantly lower migration compared to the transfection reagent-only control cells which on average migrated 11% of the original scratch width into the gap. Interestingly the absolute and percentage gap reduction was also significantly lowered by the addition of transfection reagent on its own. This is in line with the observation that the transfection process itself resulted in considerably more cell death in the TKD SW620 cell line compared to the TKD 1 and TKD 2 SW480 cells (data not shown).

It should be noted that the increased mRNA levels of the *TETs* following their overexpression were not reflected in the protein levels of any of the conditions in the three cell lines (Supplementary Figure 11). This was an unexpected observation which cannot be easily explained and is possibly a major limitation of these experiments. It is tempting to speculate that the antibodies used in this project did not bind to the overexpressed proteins. Due to the time constraints of the project and limited sample availability, however, no further tests to clarify this issue were performed.


A
0 hours
48 hours
Control
Fugene
TET1 o-e
TET2 o-e
TET3 o-e

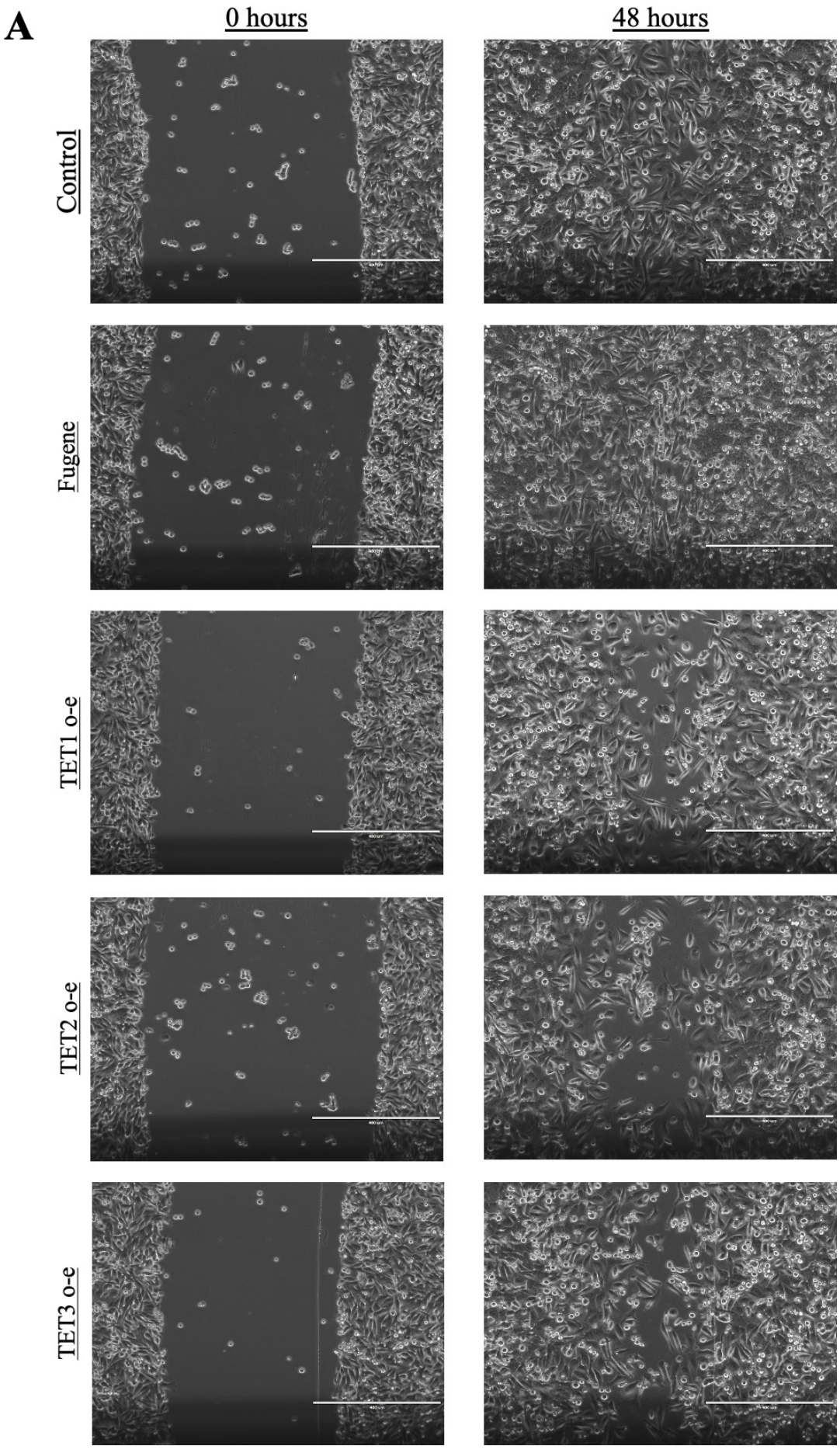


Sample comparison (percentage reduction in gap width)	Significance level
Control vs Fugene	ns
Control vs TET1 o-e	**
Control vs TET2 o-e	ns
Control vs TET3 o-e	**
Fugene vs TET1 o-e	ns
Fugene vs TET2 o-e	ns
Fugene vs TET3 o-e	**
TET1 o-e vs TET2 o-e	ns
TET1 o-e vs TET3 o-e	**
TET2 o-e vs TET3 o-e	**

Figure 5.9 The results of TET enzymes overexpression on migration of the TKD 1 SW480 cell line

(A) The representative images of migration of untransfected control, transfection reagent-only control (Fugene) and TET1/2/3 overexpressing TKD 1 SW480 cells. The images were taken at 0 hours and 48 hours after the scratch was made, at 10X magnification. Scale bar = 400 µm.

(B) The depiction of the percentage gap closure between the 0 hour and 48 hour time points assessed using the ImageJ software in untransfected control, Fugene and TET1/2/3 overexpressing TKD 1 SW480 cells. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (** $p<0.01$). The results of the statistical test are shown in the table next to the graph.


A
0 hours
48 hours
Control
Fugene
TET1 o-e
TET2 o-e
TET3 o-e

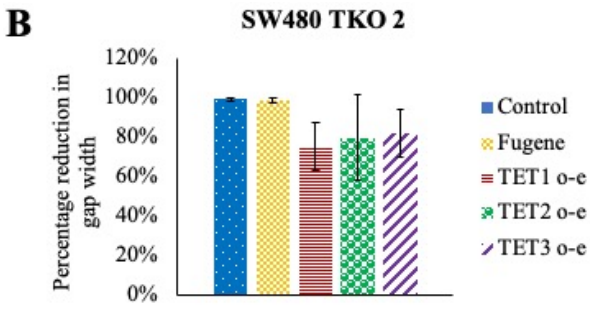
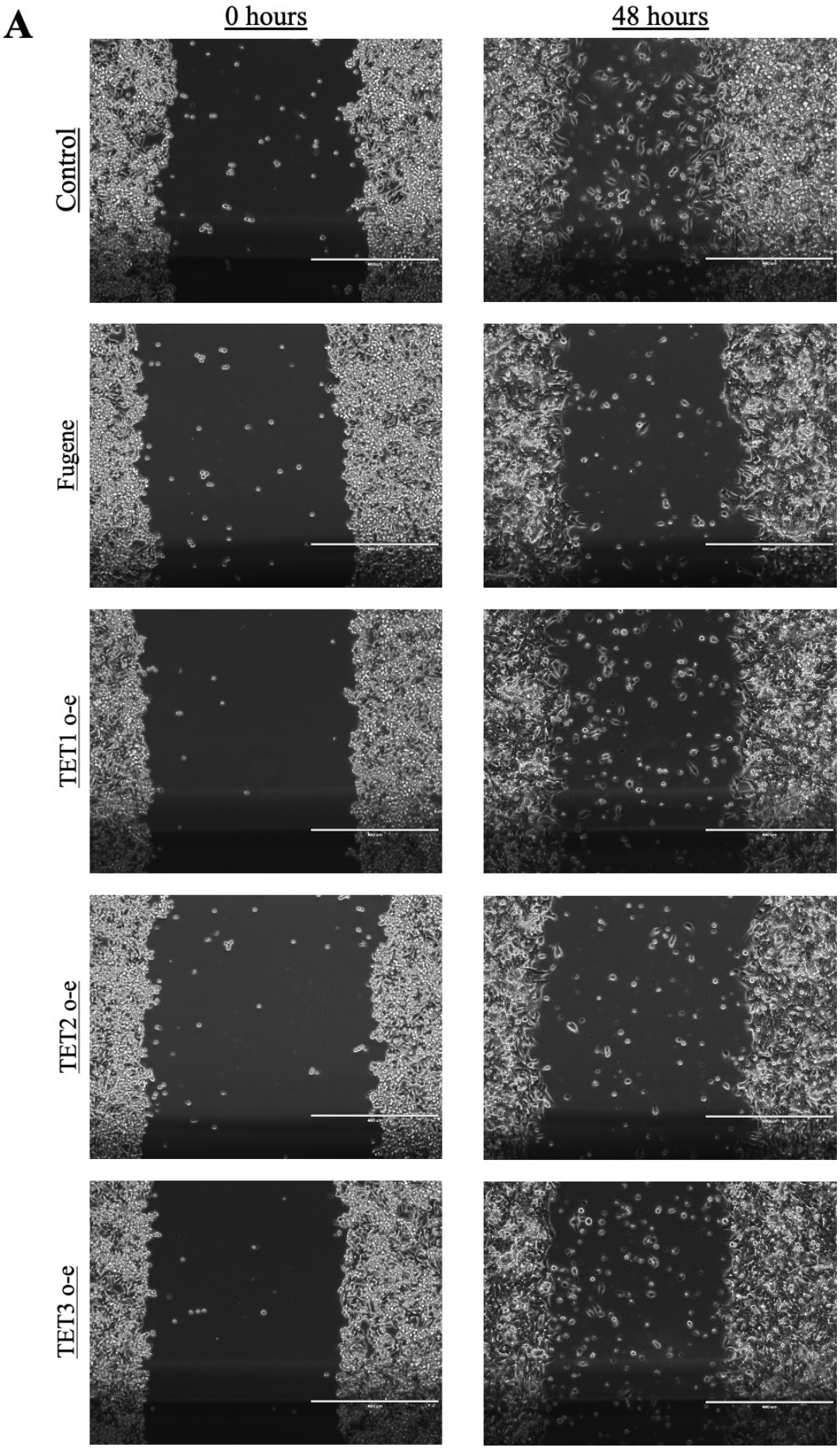
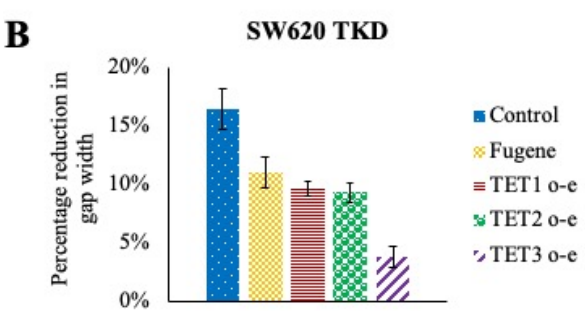


Figure 5.10 The results of TET enzymes overexpression on migration of the TKD 2 SW480 cell line

(A) The representative images of migration of untransfected control, transfection reagent-only control (Fugene) and TET1/2/3 overexpressing TKD 2 SW480 cells. The images were taken at 0 hours and 48 hours after the scratch was made, at 10X magnification. Scale bar = 400 µm.

(B) The depiction of the percentage gap closure between the 0 hour and 48 hour time points assessed using the ImageJ software in untransfected control, Fugene and TET1/2/3 overexpressing TKD 2 SW480 cells. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison.


A
0 hours
48 hours
Control
Fugene
TET1 o-e
TET2 o-e
TET3 o-e



Sample comparison (percentage reduction in gap width)	Significance level
Control vs Fugene	**
Control vs TET1 o-e	**
Control vs TET2 o-e	**
Control vs TET3 o-e	**
Fugene vs TET1 o-e	ns
Fugene vs TET2 o-e	ns
Fugene vs TET3 o-e	**
TET1 o-e vs TET2 o-e	ns
TET1 o-e vs TET3 o-e	**
TET2 o-e vs TET3 o-e	**

Figure 5.11 The results of TET enzymes overexpression on migration of the TKD SW620 cell line

(A) The representative images of migration of untransfected control, transfection reagent-only control (Fugene) and TET1/2/3 overexpressing TKD SW620 cells. The images were taken at 0 hours and 48 hours after the scratch was made, at 10X magnification. Scale bar = 400 µm.

(B) The depiction of the percentage and absolute gap closure between the 0 hour and 48 hour time points assessed using the ImageJ software in untransfected control, Fugene and TET1/2/3 overexpressing TKD SW620 cells. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (** $p<0.01$). The results of the statistical test are shown in the table next to the graph.

5.5 Discussion

This chapter provides a summary of the investigation into the effect of the manipulation of TET enzymes levels on the *in vitro* properties of two CRC cell lines - SW480 and SW620. However, these results come with the caveats of experimental limitations. The first limitation is that the cell lines obtained from Horizon Discovery had varying success with regard to the efficiency of the *TET* knockdown and still retained some evidence of enzymatic activity as indicated by the detectable persistence of low 5hmC levels in both SW480 and SW620 knockdown cell lines despite all three *TET* genes being successfully targeted at the genomic level. To date, there is no evidence for a fourth TET enzyme or a mechanism whereby 5hmC is maintained during replication in the absence of functional TET enzymes. Moreover, *TET* TKO mouse embryonic stem cells lack detectable levels of 5hmC in their genome (Dawlaty et al., 2014). In light of this evidence, we hence regard these cell lines as catalytic knockdowns rather than knockouts.

At the transcript level, the qRT-PCR results obtained indicated a greater degree of knockdown in SW480 compared to SW620 cells. None of the cell lines, however, appeared to be knockouts of the TET enzymes on mRNA level. The CRISPR/Cas9 strategy to introduce insertion/deletion mutations still allows a gene to be transcribed despite a disrupted DNA sequence and a premature STOP codon which terminates translation prematurely leading to a shortened, often non-functional protein. However, it is known that in most cases the mRNA produced is degraded via nonsense mediated decay in the nucleus before being exported out (Shi et al., 2015; Tuladhar et al., 2019). Additionally, while one of the alleles may be a functional knockout, the other can compensate for it in terms of the gene expression (Hildebrandt et al., 2015). Lastly, it is also known that a large proportion of CRISPR/Cas9 induced mutations results in a variable level of residual protein expression (Smits et al., 2019). The similar pattern of knockdown of the TET enzymes in all four SW480 knockdown cell lines may be due to TET2, which is the highest abundance TET enzyme in these cells (see section 4.4.1, Figure 4.2 B), being involved in control of the expression of *TET1* and *TET3* genes by, for example, demethylating their promoters. It should be noted that this is purely a speculation as no evidence of such regulatory mechanism has been published to date. If that is the case, the knockdown of *TET3* and *TET1* in DKD and TKD cell lines,

respectively would have little effect on their mRNA expression. In SW620 cells, the supposed *TET2* knockout had no effect on the expression of the three TET enzymes. *TET3* knockdown in the DKD cell line reduced its expression while elevating levels of *TET1* and *TET2*. None of these changes were found to be significant. The *TET1* knockdown in the TKD cells led to a significant reduction in only the *TET1* transcript relative to SKD and DKD cell lines.

The protein and 5hmC levels in the *TET* knockdown SW480 cell lines do not support qRT-PCR data showing an equal knockdown of all three *TET* transcripts in SKD, DKD, TKD 1 and TKD 2 cell lines. This could potentially be due the levels of TET proteins remaining relatively stable despite fluctuations in mRNA content. Given that TET2 is the most abundant of the TET proteins in SW480 cells, the changes in *TET1* and *TET3* transcription observed in SKD cell line might not affect the protein levels to the same extent as *TET2*. A non-significant reduction in TET protein content was observed in the SKD and DKD cell lines together with 40% lower global 5hmC levels compared to the wild type cells. The reduction in both 5hmC and TET protein levels was greater in the two TKO cell lines compared to SKD and DKD cells, which further strengthens the results of functional assays where the greatest changes were observed in the two TKD cell lines.

In SW620 cells, the different results of *TET* expression and TET protein levels analysis were also observed. The significant reduction in *TET1* content observed in TKD SW620 cells did not translate into a reduction of its protein levels. Despite this, the reduction in 5hmC levels indicates that the activity of the TET proteins is lower in the TKD cell line than the wild type counterpart. This could be due to the levels of TET1 and TET2 sufficiently reduced to elicit changes in global 5hmC levels despite the unexpected increase in TET3 content. The reduction in both *TET* transcripts and 5hmC levels was greater in the TKD SW620 cell line compared to the SKD and DKD cell lines.

The *TET* TKD thus appears to be more efficient in SW480 cells than the metastatic counterpart as seen by the greater reduction in both *TET* expression and 5hmC content. A partial limitation of this study is the absence of the mass spectrometry 5hmC content data due to its levels being below the quantification limit. This could be overcome in the future by attempting to analyse a larger quantity of the TKD DNA on the mass spectrometer to increase the 5hmC amount above that limit.

Taken together, these results indicate that the SW480 and SW620 knockdown cell lines are not complete knockouts on functional level. However, the knockdown cell lines described here represent a viable, albeit limited, model to study the effects of TET and 5hmC levels reduction in the *in vitro* CRC model. This is because of the opposite roles of the TET enzymes in cancer development reported in the literature and the scarcity of data regarding their function in CRC progression and metastasis.

It is important to note the limitations of use of the CRISPR/Cas9 gene editing system to produce clonal knockout cell lines. The process of obtaining a single clone harbouring the desired deletion from a population of transfected cells represents a massive selection bottleneck. The resulting cell line is a homogenous population of cells arising from a single cell which previously was a tiny part of a heterogenous cell line. As such, this particular cell might not be representative of the whole original population and have properties that made it successful during the selection process. This highlights the need to analyse multiple clones harbouring the same deletion before any conclusions can be made about the observed phenotype. The practice of analysing multiple clones isolated from the original transfected population will also minimise the chances that off-target effects of the CRISPR/Cas9 system play a role in any phenotypes observed with a particular clone and thus strengthens any subsequent findings. It is also important to include a non-targeting control guide RNA in the CRISPR experiment. The sequence of this guide is often a scrambled sequence of the guide targeting the genomic region of interest. This approach thus provides a further control for any off-target effects potentially arising following the CRISPR/Cas9 vector transfection into cells.

The proliferation and colony formation assay results indicate that reduced TET levels result in lower rates of proliferation, although the effects appear to be more pronounced with regards to the colony formation ability. In SW480 cells, the different rates of proliferation in the two TKD cell lines are certainly puzzling given their similar levels of *TET* knockdowns and 5hmC levels. It may reflect differences in off-target effects or properties of the single parent cell between these two cell lines, highlighting the need to analyse multiple CRISPR/Cas9 clones alluded to above, and ultimately to do full sequence analysis. TET1 was previously shown to play a role in regulating proliferation and migration in CRC cell lines *in vitro* (Tian et al., 2017) and may hence help to explain the disparity in the results if there is a difference in the degree of functional *TET1* knockdown in the two TKD SW480 cell lines which was to an extent observed on

protein level. The colony formation assay results further support this with TKD 1 cells showing no change in the number of colonies formed and an increased colony size, while TKD 2 cell line displayed a reduction in colony count compared to the wild type cells. Contrary to the proliferation assay, SKD and DKD SW480 also showed a reduction in proliferative capabilities. This is possibly due to the colony formation assay highlighting the ability of a single cell in a population to proliferate independently and establish a new colony. When the colony sizes are compared, little difference is seen between these cell lines and wild type cells. While the *TET* knockdown SW620 cells show no change in growth rate measured by the proliferation assay, the SKD, DKD and TKD cell lines have a significantly lower rate of colony formation compared to the wild type cells. In a similar fashion to the SW480 knockdown cell lines, it is hence likely that the ability of individual cells to form independent colonies is affected while there is little effect on the rate of growth of the whole cell population. Together, these results demonstrate that cells with lower TET and 5hmC levels have a modest growth reduction as would be expected given the pro-oncogenic role of TETs previously described in the literature (Tian et al., 2017; Good et al., 2018). These results are, however, counter-intuitive given the abundance of data supporting the tumour suppressive functions of the TET enzymes (see section 5.2 for details). In general, the findings described in this chapter do support the notion that TETs may play opposite roles in different cancer types.

The increased motility and ability to migrate of cancer cells is a crucial step in the metastatic cascade. The *in vitro* migration of the *TET* knockdown SW480 and SW620 cell lines was therefore directly compared using the scratch assay. In the primary SW480 cells, all the *TET* knockouts resulted in no change in migration rate compared to the wild type cell line. Interestingly, the *TET* TKD in the metastatic SW620 cell line showed little effect on proliferation but had a suppressive effect on migration of these cells (Figures 5.5 and 5.7, respectively), with the TKD cell lines showing a marked reduction in migration. This result is similar to that of Tian and colleagues who also found that *TET1* knockdown in SW620 cells decreased their invasiveness (Tian et al., 2017). These findings highlight the intriguing potential switch in the effects of the TET enzymes on cancer hallmarks depending on the stage of cancer progression (Tian et al., 2017).

This differential response of the two CRC cell lines, SW480 and SW620, to the *TET*

knockdowns in terms of their migration warranted further investigation. The initial question asked was whether a rescue in expression in any individual TET enzyme will be sufficient to reverse the phenotype in case of SW620 cells and whether TET overexpression will lead to a migratory phenotype in the SW480 cell line. In the TKD 1 and 2 SW480 cell lines, overexpression of TET1 and TET3 resulted in the greatest reduction of the migration rate. The degree of this effect was positively correlated to the degree of overexpression achieved in the TKD SW480 cells. Greater overexpression of TET3 lead to more significant reduction of the migratory phenotype in TKD 1 SW480 cells. The overexpression of TET2, meanwhile, did not result in significant changes in migration rate of neither of the TKD SW480 cell lines. This could possibly be due to the TET2 overexpression vector containing a mouse gene rather than a human one which could lead to differences in the genomic regions that gain 5hmC following its overexpression.

In the TKD SW620 cell line, the overexpression of the TET enzymes unexpectedly resulted in a further reduction of the migration rate. However, it should be noted that unlike in TKD SW480 cells, the transfection process itself significantly reduced the migration of this cell line. This is likely due to the observed high quantity of cells dying following transfection (data not shown). It is therefore possible that transfection with the overexpression plasmids resulted in an even higher death rate and hence a reduction in the distance migrated by the cells. Additionally, the overexpression of the three TET enzymes achieved was lower in the TKD SW620 cells compared to the TKD 1 and TKD 2 SW480 cell lines. Given the caveat that the overexpression rescue experiment resulted in a higher than physiologically normal transcription, further rescue experiments with catalytically inactive mutant TET enzymes are required.

Time constraints of this project did not allow experiments to be performed that would further elucidate the role of the TET proteins in this *in vitro* CRC metastasis model. Assessment of the invasion of TET knockdown SW480 and SW620 cell lines through the use of transwell assays, previously utilised to assess the effect of TET levels manipulation on cancer cell invasion (Tian et al., 2017; Yu et al., 2020), would complement the migration data and provide further insight into the role of these enzymes in the metastatic process. A generation of SW480 and SW620 cell lines stably overexpressing each individual TET enzyme as well as all three TET proteins at once would allow to assess the effect of increased TET and 5hmC levels on *in vitro*

proliferation and migration. Lastly, tagging the aforementioned cell lines with fluorescent protein and luciferase reporters and introducing them into an animal model such as nude mice would provide information about how the *in vitro* findings described in this chapter translate to tumour growth and metastasis *in vivo*.

Chapter 6

The changes in 5hmC and TET enzyme levels during cellular differentiation

6.1 Chapter summary

The aim of this study was to investigate the changes in *TET* expression and 5hmC content during the process of cellular differentiation and whether this knowledge can be used to improve the *in vitro* differentiation process.

The data describing the link between dynamic changes in 5hmC distribution and gene expression in mouse small intestine was further validated following controversy regarding the role of TET1 in mouse intestinal differentiation. A differentiation protocol of hESCs to hepatocyte-like cells (HLCs) was then used to provide a novel insight into fluctuation in *TET* and 5hmC levels during *in vitro* generation of liver cells. The *TET1* transcript was found to be the most abundant of the *TETs* in H9 ESCs and its levels gradually decreased during differentiation while the opposite pattern was seen for *TET2* and *TET3*. The global 5hmC content displayed a "U" pattern, initially dropping as the cells differentiated towards definitive endoderm (DE) and increased again in HLCs to levels higher than those observed in ESCs. A transient shRNA-mediated knockdown of *TET1* levels in ESCs was utilised in an unsuccessful attempt to improve differentiation to the DE stage.

Taken together these results provide a novel insight into the changes in 5hmC and TET levels during mouse intestinal differentiation and *in vitro* generation on HLCs from hESCs. They also highlight a potential strategy for improving the differentiation of HLCs from pluripotent sources by altering the levels of TET proteins and 5hmC distribution despite an inconclusive first attempt at achieving such improvement.

6.2 Introduction

The process of cellular differentiation involves a complex myriad of chromatin remodelling and gene expression changes that allow cells containing the same genome to change their identity from a pluripotent progenitor to the differentiated progeny. Tumorigenesis and metastasis occur when the processes of cellular differentiation go awry. Mouse embryonic stem cells (mESCs) have been extensively used to study the role of the TET enzymes in pluripotency and differentiation (comprehensively reviewed by Ross and Bogdanovic, 2019). *Tet3* is known to be the most abundantly expressed *Tet* transcript in the mouse oocyte and zygote with *Tet1* and *Tet2* mRNA detected at very low levels (Iqbal et al., 2011; Wossidlo et al., 2011). However, as the cells differentiate towards the blastocyst stage, from which ESCs are derived, the pattern of *Tet* expression changes starkly. *Tet1* is the most highly expressed *Tet* transcript in the mESC while *Tet2* levels are few-fold lower but still easily detected and *Tet3* is barely detectable in these cells (Koh et al., 2011). Similarly, the levels of *Tet1* and *Tet2* are elevated following generation of iPCSs from mouse embryonic fibroblasts, while the abundance of the *Tet3* transcript is reduced in the process (Koh et al., 2011). Both mESCs and iPSCs are also known to have high 5hmC levels compared to more differentiated, multipotent cells (Szwagierczak et al., 2010; Koh et al., 2011). While the *Tet* triple knockout mESCs are able to maintain their pluripotency and normal morphology (Lu et al., 2014), their differentiation to embryonic bodies (Dawlaty et al., 2014) and neural lineages (Li et al., 2016) is severely impaired. Additionally, *Tet1/Tet3* double and *Tet* triple knockout embryos show numerous developmental defects and embryonic lethality (Kang et al., 2015; Dai et al., 2016; Li et al., 2016).

The changes in *TET* and 5hmC levels have also been studied during hESC differentiation. *TET* TKO hESCs, similarly to mESCs, do not show any defects in proliferation, morphology or pluripotency maintenance but display differentiation impairments (Verma et al., 2018). Differentiation of H9 hESCs towards the haematopoietic lineage was found to be associated with a large reduction of *TET1* and gradual upregulation of *TET2*, with a small increase in *TET3* transcript levels also noted (Langlois et al., 2014). *TET2* was additionally described to be critical for haematopoietic differentiation with a *TET2* knockdown skewing H9 ESCs differentiation from mesoderm to neuroectoderm (Langlois et al., 2014). Li and

colleagues provided an insight into changes in *TET* and global 5hmC levels during a multi-stage H1 hESCs differentiation protocol towards the pancreatic lineage (Li et al., 2018). During the differentiation protocol, global 5hmC content was observed to dip initially during the first two stages by approximately 70% followed by a recovery to approximately 80% of the levels seen in hESCs in the last differentiation stage, the pancreatic endoderm (Li et al., 2018). Similarly to the observations made by Langlois and colleagues, *TET1* levels were found to decrease during the differentiation process, while *TET2* and *TET3* transcripts increased their abundance as the cells differentiated towards the pancreatic endoderm (Langlois et al., 2014; Li et al., 2018).

6.2.1 The role of the TET enzymes and 5hmC in mouse intestinal differentiation

The mammalian intestine and colon are an easily accessible model system in which to study the epigenetics of multipotent stem cells and their differentiated progeny (Sheaffer et al., 2014; Van Der Heijden and Vermeulen, 2019). The epithelium of mammalian intestine is a series of villi projections that grow from crypts at their base and protrude into the intestinal lumen (see section 1.2.4, Figure 1.2). The crypts contain the intestinal stem cells (ISCs) and Paneth cells while the villi are formed of fully differentiated cells responsible for nutrient absorption (Beumer and Clevers, 2016). It was shown in both mouse and human that while global 5mC levels remain relatively constant across the intestinal crypt and villus, the global 5hmC content is high in the nonproliferating, differentiated villus progeny and low in the rapidly proliferating stem cell compartment (Uribe-Lewis et al., 2015; Kim et al., 2016; Uribe-Lewis et al., 2020). The role and abundance of the TET enzymes during the intestinal differentiation, however, remains controversial. Kim and colleagues demonstrated a higher expression of *Tet1* in the mouse ISCs compared to the other two *Tets* and highlighted its essential role in maintenance and function of the ISCs population in the mouse small intestine (Kim et al., 2016). Additionally, they showed that the expression of *Tet1* was significantly downregulated, and the levels of *Tet2* and *Tet3* were significantly upregulated, in the differentiated villus cells compared to the ISCs (Kim et al., 2016). On the other hand, a recently published study from the Murrell group (Uribe-Lewis et al., 2020) found that the levels of *Tet1* are low in the ISCs population and decrease further during intestinal differentiation. This is consistent with the previous report by

the same group showing low *TET1* levels compared to *TET2* and *TET3* expression in the human colon (Uribe-Lewis et al., 2015). Additionally, the expression of *Tet2* and *Tet3* displayed a very modest upregulation in the differentiated progeny and thus cannot account for the increase in global 5hmC levels (Uribe-Lewis et al., 2020). Lastly, Chapman and colleagues demonstrated an increase in *TET1* expression during *in vitro* differentiation of colonocytes, which is in direct contrast to the two studies described above (Chapman et al., 2015). Given this discrepancy in results, the expression and role of the TET enzymes in intestinal differentiation warrants further investigation.

6.3 Experimental objectives

Aim: to examine the changes in 5hmC and *TET* levels during adult and embryonic stem cell differentiation of the small intestine and the liver to further our understanding of their role in untransformed tissue.

Objectives:

1. Confirm the changes in *TET* expression during mouse small intestine differentiation
2. Establish the changes in *TET* and global 5hmC levels during the differentiation of human embryonic stem cells to hepatocyte-like cells
3. Attempt to improve the differentiation efficiency of human embryonic stem cells to hepatocyte-like cells through a *TET1* knockdown

6.4 Results

6.4.1 The confirmation of changes in *TET* expression during mouse small intestine differentiation

The Murrell group previously generated RNA-seq and hmeDIP-seq data for two cell types in the mouse small intestine: the stem cell crypt progenitors (CD24a_Mid cells) and the differentiated villus progeny (CD24_Neg cells) (Uribe-Lewis et al., 2020). The aforementioned discrepancy in *Tet* levels observed by the Murrell group and Kim and colleagues meant that further confirmation of changes in *Tet* expression during mouse intestinal differentiation was required (Kim et al., 2016). Thus, more mouse ISCs and differentiated intestinal progeny cells were obtained and the *Tet* transcription levels was independently assessed by qRT-PCR (Figure 6.1). This independent qRT-PCR validation of the *Tet* expression showed a similar pattern in changes of *Tet* abundance during the differentiation of mouse small intestine as observed by RNA-seq (Figure 6.1 A). The levels of *Tet1* were reduced by 81% (compared to a 90% reduction observed by RNA-seq) while the levels of *Tet2* and *Tet3* were increased by 39% and 30%, respectively (compared to an increase of 54% and 17%, respectively, observed by RNA-seq) in the differentiated villus cells compared to crypt progenitor cells. In order to assess the levels of expression of individual *Tets* in both cell lines, their levels were normalized to the *Polr2a* reference gene, revealing that *Tet3* is the most abundant *Tet* in both cell types (Figure 6.1 B). Importantly, the expression of the three *Tets* and the *Polr2a* reference gene were adjusted to account for the primer efficiencies (see Supplementary Figure 12 for details).

The discrepancy in the expression of *Tet1* reported by Kim and co-authors and Uribe-Lewis and colleagues could be due to alternative exon usage (Kim et al., 2016; Uribe-Lewis et al., 2020). The exon targeting by the *Tet1* qRT-PCR primer pairs used by both groups was hence directly compared. The mouse *Tet1* gene contains thirteen exons (Towers et al., 2018). The *Tet1* primer pair used by Kim and colleagues (Kim et al., 2016) targets the exon-exon junction between exons 12 and 13 while the primer pair used to generate the results presented in Figure 6.1 A and B binds to a sequence within exon 4 (data not shown). Two isoforms of *Tet1* were previously described in embryonic and adult mouse tissues (Zhang et al., 2016). The full-length isoform was found to be expressed in mESCs and PGCs and absent in adult tissues, while the expression of the

shorter isoform lacking the first two exons was progressively increased during neuronal differentiation of mESCs. There is no evidence, however, of the alternative use of the exons targeted by the two primer pairs mentioned above. It is therefore unlikely that different exon usage contributes to the reported differences in *Tet1* levels in mouse ISCs.

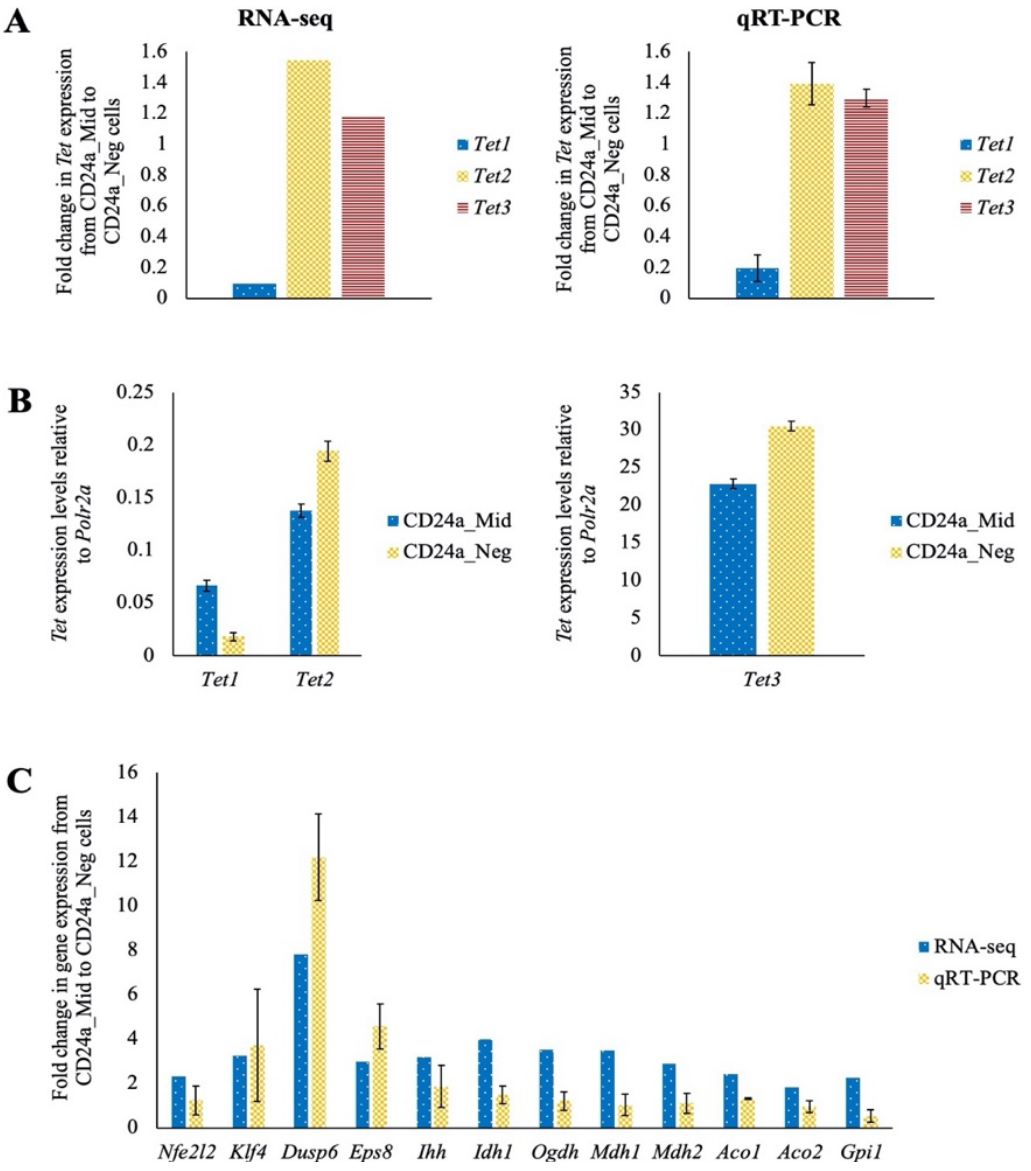


Figure 6.1 The validation of gene expression changes occurring during differentiation of mouse stem cell colonic crypt progenitors to differentiated villus progeny

(A) and (B) The comparison of fold change in expression and gene expression relative to the reference gene, respectively, of the *Tet* transcripts between the mouse colonic crypt progenitors (CD24a_Mid cells) and differentiated villus progeny (CD24a_Neg cells) obtained by RNA-seq (Uribe-Lewis et al., 2020) and qRT-PCR. The *Tet* mRNA levels were normalised to *Polr2a* reference gene. Both graphs represent the results of analysis of three independent biological replicates and the error bars display the standard deviation of three independent replicates.

(C) The comparison of fold change in expression of the genes selected for validation between CD24a_Mid and CD24a_Neg cells obtained by RNA-seq and qRT-PCR. The mRNA levels of the chosen genes were normalised to *Polr2a* reference gene. Both graphs represent the results of analysis of three independent biological replicates and the error bars display the standard deviation of three independent replicates.

6.4.2 Assessment of the changes in *TET* and global 5hmC levels during the differentiation of human embryonic stem cells to hepatocyte-like cells

The analysis of the changes in *Tet* and 5hmC levels during adult stem cell differentiation in the mouse small intestine was followed by the assessment of *TET* abundance and global 5mC and 5hmC content in hESC differentiation. The specific protocol utilised in this chapter allows the stage-wise differentiation of hESCs to DE stage, a bipotent hepatic progenitor cell termed hepatoblast (HB stage) and finally to the HLC stage. This differentiation path was chosen in an attempt to model the influence of the TET proteins and 5hmC on normal liver development. This is important to understand in the light of the preliminary results (section 4.2) from the Murrell lab suggesting that liver metastasis may require increased TET activity and upregulation of global 5hmC.

The H9 ESCs display a number of morphological changes during differentiation to HLCs (Figure 6.2). The elongated, spiky morphology of the H9 ESCs disappeared at the DE stage giving way to a more compact, triangular and polygonal cell shape. This step was also accompanied by a large amount of cell death as visualised by the smaller and brighter dead cells floating above the DE monolayer in Figure 6.2. HBs were defined by a further transition to a fully polygonal cell shape, forming a mosaic-like monolayer with defined boundaries between cells. They progressed to HLCs, which were characterised by a high cytoplasmic to nuclear ratio with smaller, well defined nuclei and occasional binucleated cells as well as the presence of multiple vacuoles which is in line with previous observations of hESC-derived and human liver-derived hepatocytes (Hay et al., 2008; Yu et al., 2012; Bartlett et al., 2014).

The successful differentiation of the H9 ESCs was confirmed by qRT-PCR (Figure 6.3). The levels of *OCT4* and *ALB* transcripts which are known to decrease (Hay et al., 2008) and increase (Yu et al., 2012) during the hepatic differentiation process, respectively, were measured at each stage of the protocol. The *ALB* content did not change at the DE stage and remained at low levels of expression (Figure 6.3 A). Upon reaching the bipotent HB stage, the *ALB* levels increased by almost 200-fold compared to the H9 ESCs, followed by approximately 34,000-fold elevation in expression in the HLCs.

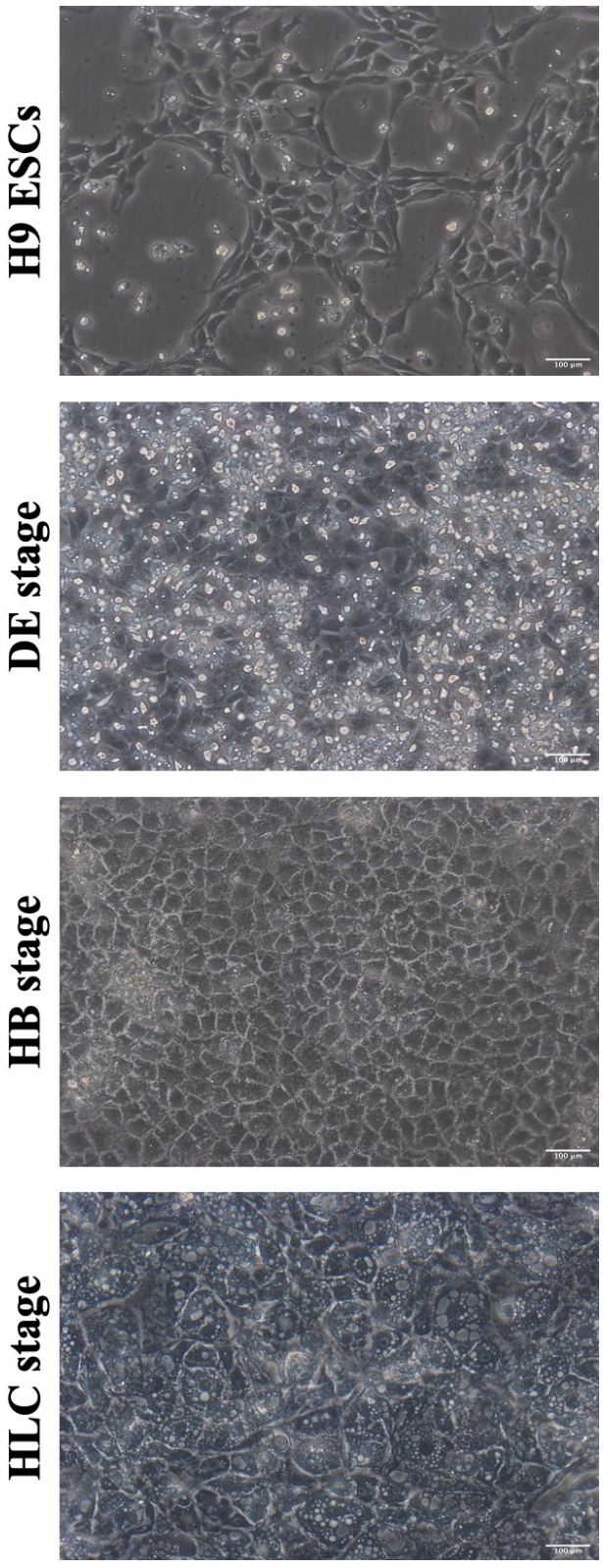


Figure 6.2 The morphological changes during human embryonic stem cell differentiation to hepatocyte-like cells

The representative images of human H9 embryonic stem cells (H9 ESCs), definitive endoderm cells (DE stage), hepatoblast cells (HB stage) and hepatocyte-like cells (HLC stage) taken at 20X magnification. Scale bar = 100 μm.

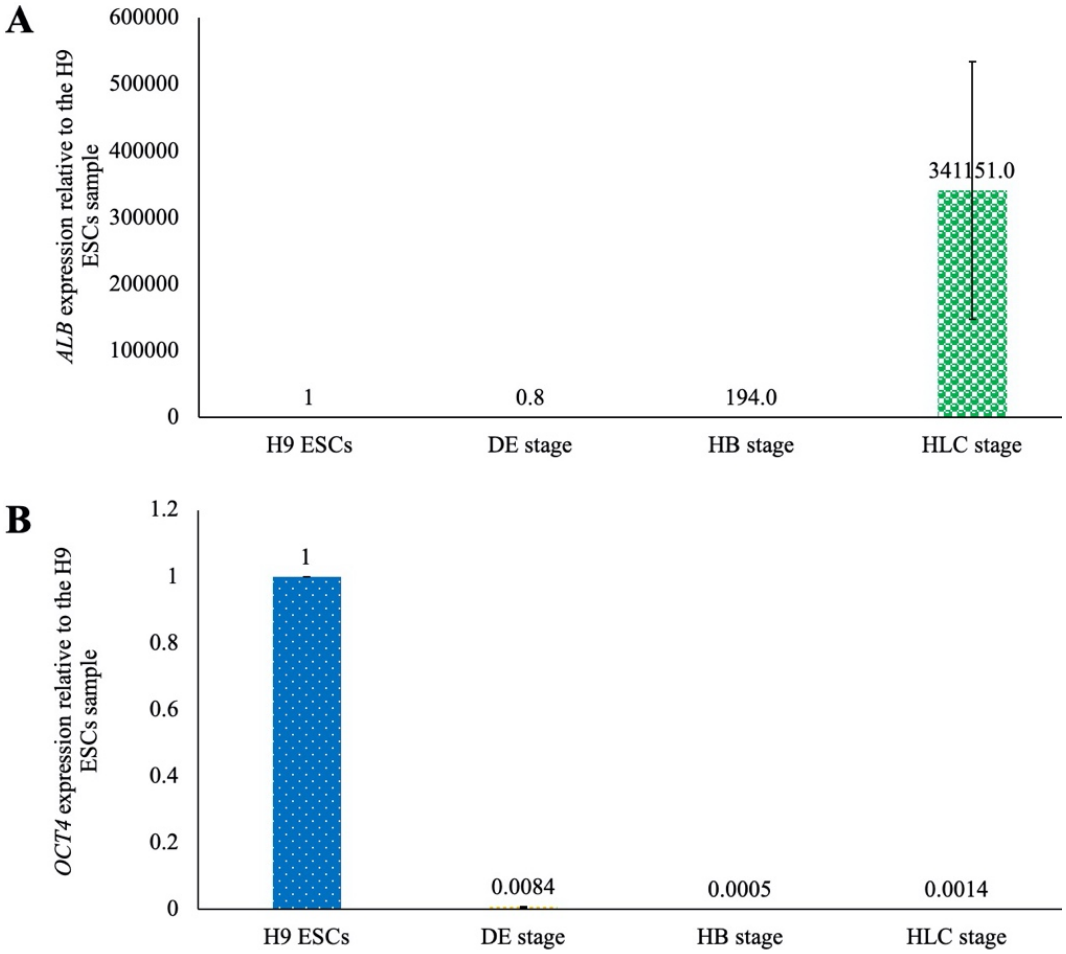


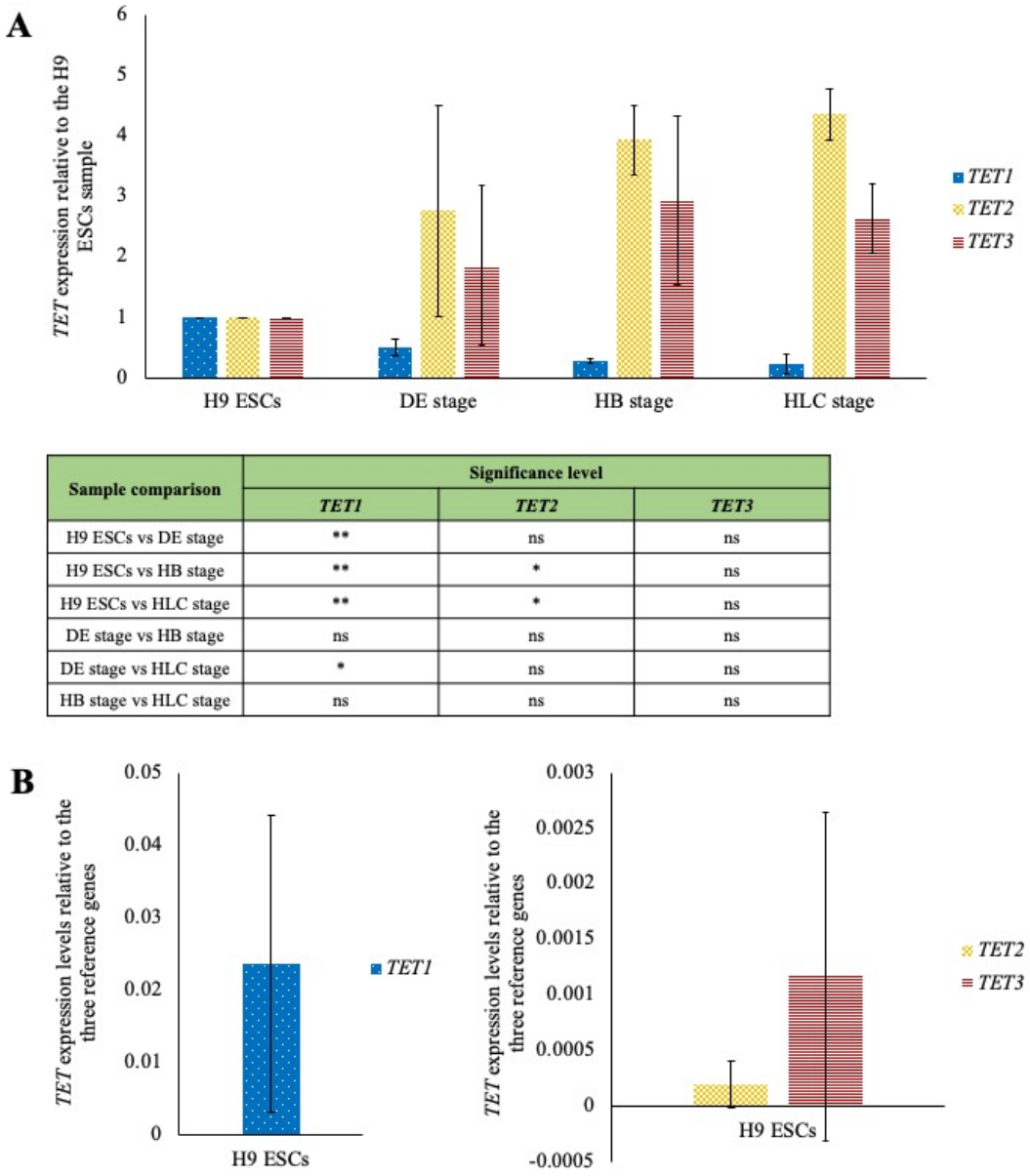
Figure 6.3 Confirmation of successful H9 human embryonic stem cell differentiation to hepatocyte-like cells
The confirmation of (A) *ALB* and (B) *OCT4* expression changes during embryonic stem cell differentiation to hepatocyte-like cells. H9 ESCs = human H9 embryonic stem cells, DE = definitive endoderm, HB = hepatoblast, HLC = hepatocyte-like cell. The *ALB* and *OCT4* mRNA levels were normalised to the average expression of *RPL13A*, *GAPDH* and *ACTB* reference genes as previously described (Baxter et al., 2015). The error bars represent the standard deviation of three independent replicates.

The TET enzymes have previously been described to play critical roles in the differentiation of ESCs (Langlois et al., 2014; Li et al., 2018; Verma et al., 2018) and the formation of iPSCs through a dedifferentiation process (Gao et al., 2013; Zhu et al., 2014; Secardin et al., 2020). Based on this research, it was hypothesised that the abundance of the individual TETs as well as 5hmC will fluctuate at different stages of the H9 ESC differentiation protocol to HLCs. The levels of the three *TET* transcripts were hence assessed by qRT-PCR in the same RNA samples from the four cell types as analysed above (Figure 6.4). The expression of *TET1* was found to be gradually and

significantly reduced at all stages of the differentiation protocol: the DE cells had a fold change of 0.5 compared to the *TET1* H9 ESCs levels, and this value decreased further to 0.28 and 0.23 relative to the stem cell expression in HBs and HLCs, respectively (Figure 6.4 A). Conversely, the expression of *TET2* and *TET3* gradually increased during the differentiation process. *TET2* levels increased 2.76-fold at the DE stage, followed by an incremental 3.94- and 4.36-fold elevation in HBs and HLCs, respectively, relative to *TET2* expression in H9 ESCs. The increase in *TET2* levels in the latter two stages was found to be highly significant. *TET3* transcript content was elevated 1.86-fold in the DE cells, followed by 2.94-fold and 2.64-fold increase in HBs and HLCs compared to *TET3* levels in H9 ESCs. However, none of these changes were found to be significant.

The expression of the individual *TETs* was assessed in the H9 ESCs to better understand which of the three enzymes is contributing to the maintenance of 5hmC in these cells and to establish a baseline for the changes in *TET* expression shown in Figure 6.4 A (Figure 6.4 B). *TET1* was found to have the highest expression out of the three enzymes, approximately 122-fold and 20-fold higher than *TET2* and *TET3*, respectively Intriguingly, the higher absolute values of *TET1* expression in in the H9 ESCs suggest that despite the changes in *TET* expression levels during the differentiation process depicted in Figure 6.4 A, *TET1* is still the most predominant *TET* transcript. It should be noted that the large standard deviation observed in the qRT-PCR experiments likely stems from varying efficiency of differentiation in the three independent experiments.

The qRT-PCR analysis was followed by the corresponding assessment of global 5mC and 5hmC levels in the same samples as analysed in Figure 6.3 and Figure 6.4 by the dot blot technique (Figure 6.5 A). The 5mC content was observed to progressively drop to roughly 74% in the HLCs compared to the initial amount observed in H9 ESCs. Global 5hmC levels were observed to reduce to 91% of the H9 ESCs content at the DE stage, followed by an increase to 120% and 116% of the stem cell levels in HBs and HLCs, respectively. The mass spectrometry analysis revealed a small drop in global 5mC levels in the DE and HLC stages of differentiation, with no difference observed at the HB stage compared to the content observed in H9 ESCs (Figure 6.5 B). None of these changes were found to be significant.



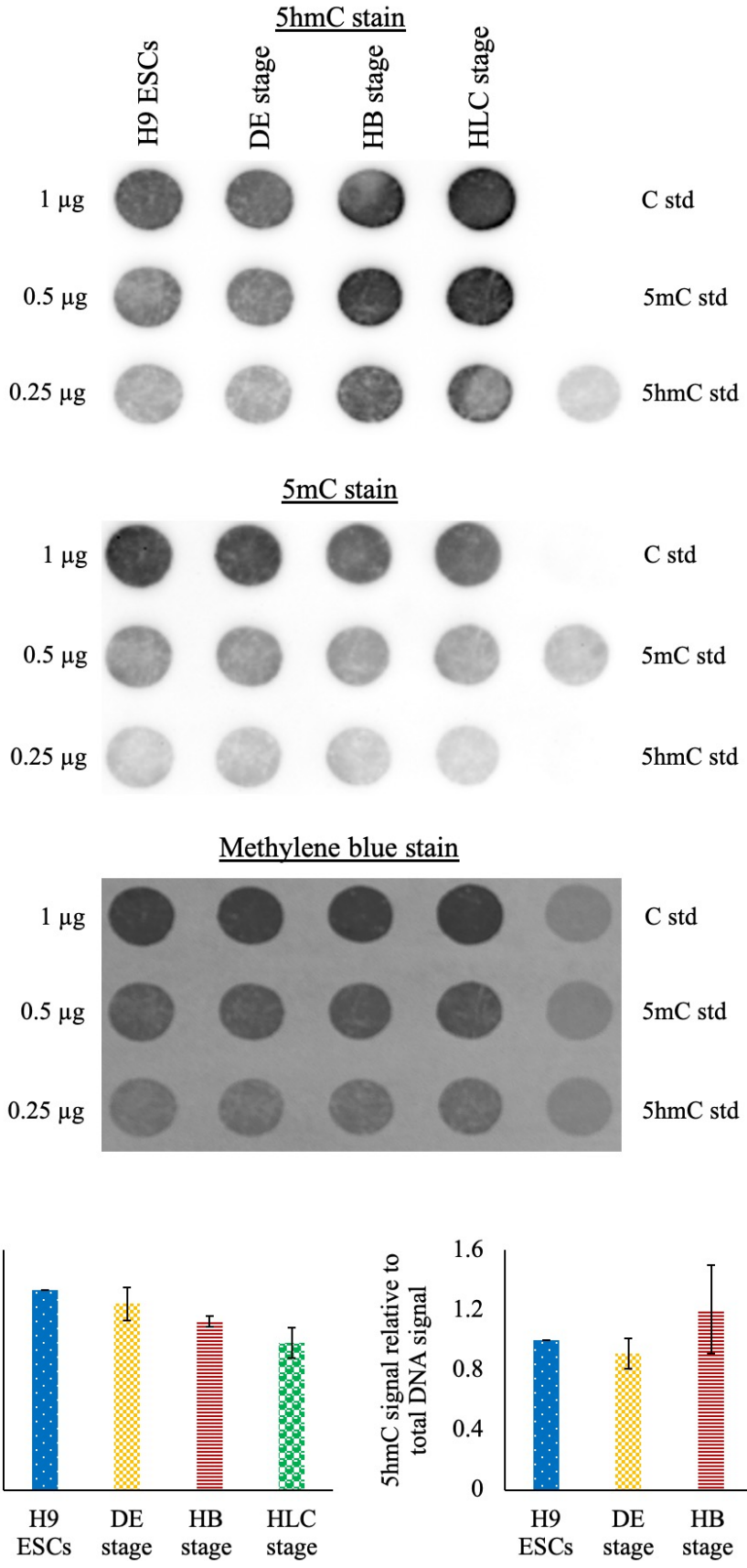
Sample comparison	Significance level		
	TET1	*TET2*	*TET3*
H9 ESCs vs DE stage	**	ns	ns
H9 ESCs vs HB stage	**	*	ns
H9 ESCs vs HLC stage	**	*	ns
DE stage vs HB stage	ns	ns	ns
DE stage vs HLC stage	*	ns	ns
HB stage vs HLC stage	ns	ns	ns

Figure 6.4 The TET enzymes significantly change their transcript levels during human embryonic stem cell differentiation to hepatocyte-like cells

(A) The changes in levels of the *TET* transcripts during embryonic stem cell differentiation to hepatocyte-like cells. H9 ESCs = human H9 embryonic stem cells, DE = definitive endoderm, HB = hepatoblast, HLC = hepatocyte-like cell. The *TET* mRNA levels were normalised to the average expression of *RPL13A*, *GAPDH* and *ACTB* reference genes as previously described (Baxter et al., 2015). The error bars represent the standard deviation of three independent differentiation experiments. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$; ** $p<0.01$; ns = not significant). The results of the statistical test are shown in the table below the graph.

(B) The expression levels of the three *TET* transcripts in H9 ESCs relative to the average expression of *RPL13A*, *GAPDH* and *ACTB* reference genes. The error bars represent the standard deviation of three independent differentiation experiments.

A


5hmC stain
H9 ESCs
DE stage
HB stage
HLC stage
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
5mC stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std
Methylene blue stain
1 µg
0.5 µg
0.25 µg
C std
5mC std
5hmC std

5mC signal relative to total DNA signal
1.2
1
0.8
0.6
0.4
0.2
0
H9 ESCs
DE stage
HB stage
HLC stage
5hmC signal relative to total DNA signal
1.6
1.2
0.8
0.4
0
H9 ESCs
DE stage
HB stage
HLC stage

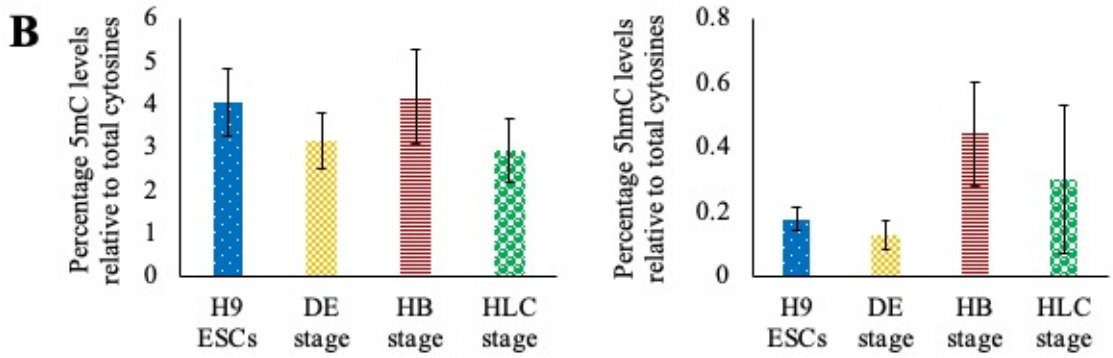


Figure 6.5 The global 5hmC content shows significant changes during human embryonic stem cell differentiation to hepatocyte-like cells
(A) The representative staining results of dot blot assessment of 5hmC and 5mC levels in human H9 embryonic stem cells (H9 ESCs), definitive endoderm cells (DE stage), hepatoblast cells (HB stage) and hepatocyte-like cells (HLC stage). The results have been normalised to the H9 ESCs sample as the replicates were analysed at different points in time. The error bars represent the standard deviation of three independent differentiation experiments.
(B) The mass spectrometry results of 5mC and 5hmC quantification in human H9 embryonic stem cells (H9 ESCs), definitive endoderm cells (DE stage), hepatoblast cells (HB stage) and hepatocyte-like cells (HLC stage). The signal obtained for each cytosine modification is displayed as a percentage of signal obtained for all cytosines. The error bars represent the standard deviation of three independent differentiation experiments. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison.

The 5hmC levels displayed a similar pattern of changes to the one observed in the dot blot experiments. There was an initial drop in 5hmC content from 0.18% of all cytosines in the stem cells to 0.13% at the DE stage. This was followed by an increase in 5hmC levels to 0.44% of all cytosines in the HB cells. The HLC were characterised by 0.30% 5hmC content. None of the above changes which were found to be significantly different to the levels found in any of the other stages.

Given the large variation observed in 5hmC and *ALB* levels in the HLCs, it was hypothesised that varying degrees is differentiation efficiency at the last stage of the protocol could be impacting the global 5hmC levels. In order to test this proposition, *ALB* levels relative to the three reference genes were compared to the 5hmC content for each of the three independent differentiation experiments (Supplementary Figure 13). No correlation between these two variables was found thus indicating that the variation in the 5hmC levels observed in HLCs is unlikely due to differences in the hepatic

phenotype achieved as measured by *ALB* expression.

These results provide a detailed description of changes in 5hmC and *TET* transcript levels during hepatic differentiation of hESCs. They also highlight the distinct roles that the individual TET enzymes are likely to play at different stages of the differentiation process.

6.4.3 Assessment of the effect of transient *TET1* knockdown on the differentiation efficiency of human embryonic stem cells to hepatocyte-like cells

The generation of HLCs from pluripotent stem cell cultures has significant promise in, among others, liver transplantation in liver failure sufferers and drug toxicity screening in pre-clinical treatment development (Hay et al., 2008; Yu et al., 2012). This is due to shortage of primary hepatocyte sources, their rapid dedifferentiation in cell culture conditions and heterogenous genetic backgrounds (Corbett and Duncan, 2019). However, HLCs are known to resemble more closely the human foetal hepatocytes rather than adult ones, in part in terms of the expression of the CYP enzymes responsible for liver metabolism which poses issues in terms of their utility in drug screening (Baxter et al., 2015). Development of strategies allowing for more efficient differentiation to hepatic cells and achieving an overall phenotype that resembles the adult hepatocyte characteristics more closely is hence urgently needed.

The high levels of *TET1* observed in the H9 ESCs compared to the other two *TET* transcripts and the large reduction in its expression upon differentiation to the DE stage led to the hypothesis that a knockdown of the *TET1* levels in the H9 ESCs prior to the start of the differentiation protocol could improve the differentiation efficiency to the DE cells. This step was also targeted with the knockdown strategy as it is only three days long and hence ideal for this proof-of-principle experiment. *TET1* was hence targeted with two different shRNA constructs in a transient transfection experiment. The degree of knockdown of the *TET1* levels following the completion of differentiation to the DE stage was subsequently assessed by qRT-PCR (Figure 6.6 A). The *TET1* shRNA 1 construct significantly reduced the *TET1* expression relative to the relative to the non-targeting shRNA against firefly luciferase mRNA (0.75-fold). The *TET1* shRNA 2 vector on the other hand did not produce a significant reduction in *TET1*

expression. The process of transfection itself led to a significant, 2.07-fold upregulation of *TET2*. The transfection with the *TET1* shRNA 1 construct led to a statistically significant reduction in *TET2* expression relative to the non-targeting control (0.38-fold). Similarly, the levels of *TET3* were also reduced in *TET1* shRNA 1 transfected samples relative to the luciferase shRNA control (0.75-fold). There was no significant change observed in *TET3* expression following the *TET1* shRNA 2 transfection, however, *TET2* levels exhibited a 0.64-fold change relative to the non-targeting luciferase shRNA control. Intriguingly, *OCT4* levels were significantly reduced to between 24 and 30% of the control expression in all transfected samples thus indicating that it is caused by the process of transfection itself.

The percentage of cells differentiated to the DE stage 72 hours after the start of the differentiation protocol was assessed by fluorescence-activated cell sorting (FACS) (Figure 6.6 B). The DE cell population was defined as positive for both CD117 (c-KIT) and CD184 (CXCR4) cell surface markers, which have been previously extensively used and validated in the literature (Holtzinger et al., 2015; Jiang et al., 2015; Mahaddalkar et al., 2020). In the untransfected control sample, only approximately 38% of cells reached the DE differentiation stage thus highlighting the inefficiency of this step (see Supplementary Figure 14 for representative FACS results). All three transfected samples showed a lower proportion of the DE cells than the untransfected control, but none of these changes were found to be significant. The *TET1* shRNA 2 transfection resulted in a higher percentage of DE cells than the non-targeting control although the difference was small, approximately 6%, and not highly statistically significant. Additionally, given that this construct did not alter the *TET1* levels (Figure 6.6 A), this is likely due to inherent variation in DE induction efficiency between independent experiments.

Taken together, these results indicate that a small, transient knockdown of *TET1* levels is not sufficient to improve the differentiation efficiency of H9 hESCs to DE.

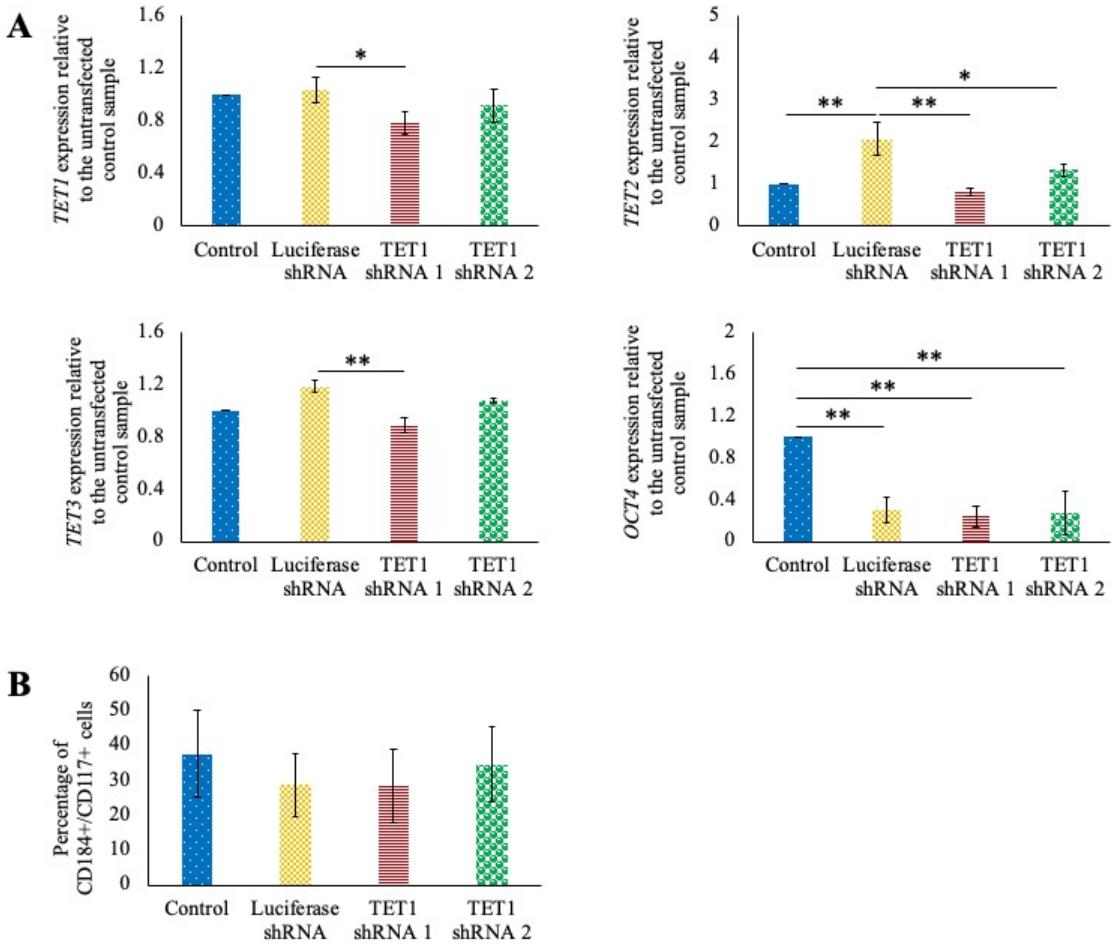


Figure 6.6 The knockdown of the *TET1* levels in human embryonic stem cells does not improve the differentiation to the definitive endoderm stage

(A) The validation of *TET1* transcript knockdown in H9 human embryonic stem cells. The *TET1*, *TET2*, *TET3* and *OCT4* mRNA levels were normalised to the average expression of *RPL13A*, *GAPDH* and *ACTB* reference genes as previously described (Baxter et al., 2015). The results were normalised to the untransfected control sample. The error bars represent the standard deviation of three independent replicates. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison (* $p<0.05$; ** $p<0.01$).

(B) The graphical representation of the percentage of cells that stained positive for definitive endoderm markers: CD184 and CD117, as measured by fluorescence-activated cell sorting. The error bars represent the standard deviation of three independent differentiation experiments. The significance was established using a one-way ANOVA followed by the Tukey's single-step multiple comparison.

6.5 Discussion

This chapter describes the investigation into the role of the TET enzymes and 5hmC during cellular differentiation of both adult and embryonic stem cells. The data previously generated, and recently published, by the Murrell lab regarding the changes in levels and distribution of 5hmC and its link to gene expression alterations during the maturation of mouse small intestine is thoroughly validated (Uribe-Lewis et al., 2020). The changes in global 5hmC levels and *TET* expression during differentiation of H9 hESCs to HLCs are also described, representing a novel contribution to the understanding of the role of these three enzymes and their product in the process of cellular differentiation. Lastly, an attempt to improve the differentiation efficiency of H9 cells to DE through a knockdown of *TET1* levels is described.

The differentiation of mouse small intestine represents a well-defined and easily accessible model for studying adult stem cell differentiation. The maturation of the colonic stem cells residing in the intestinal crypts to differentiated villus cells is accompanied by a plethora of gene expression changes facilitating the transition from self-renewal and maintenance to phenotype conducive with nutrient absorption and metabolism (Mariadason et al., 2005; Chang et al., 2008; Hiramatsu et al., 2019). Previous studies have described the changes in 5mC levels and distribution during mouse intestinal differentiation (Kaaij et al., 2013; Sheaffer et al., 2014; Kazakevych et al., 2017), thus providing insight into epigenetic regulation of this process. Given that 5hmC is thought to be broadly associated with activation of gene expression, it is unsurprising that it was also found to be largely positively correlated with changes in transcript levels during mouse intestinal differentiation (Kim et al., 2016). Kim and colleagues described *Tet1* as the most abundant *Tet* transcript in the crypt stem cells and severely disturbed crypt-villus axis morphology in *Tet1* knockout mice (Kim et al., 2016). They thus concluded that TET1 is critical for the function of mouse intestinal crypt stem cells. The data presented in Figures 6.1 A and B, however, directly contradicts this result, showing that both *Tet2* and *Tet3* are expressed at higher levels than *Tet1* in both the stem cells and differentiated progeny and that *Tet3* is the highest expressed *Tet* in both cell types. This discrepancy in results could be due to different approach to isolating the ISCs population from the mouse small intestine. Kim and colleagues utilised the high expression of *Lgr5* as a marker for ICSs while Uribe-Lewis

and co-authors isolated these cells based on their CD24a levels (Kim et al., 2016; Uribe-Lewis et al., 2020). While the *Tet1* primer pairs used by both groups target different exons, there is no evidence of alternative usage of the targeted *Tet1* exons as discussed in section 6.4.1. Alternative splicing is therefore unlikely to account for the differences reported by the two publications. Kim *et al.*, however, did not specify whether their primers have been assessed for their amplification efficiency and utilised a different reference gene (*Tbp*), which could also contribute to the difference in *Tet* expression reported (Kim et al., 2016). Additionally, the changes in *Tet2* and *Tet3* expression observed here are small, while the already low expression of *Tet1* in the ISCs is lowered further in the differentiated progeny. This result further highlights that the changes in the expression of the three *Tet* genes are unlikely to underlie the global increase in 5hmC observed during mouse intestinal differentiation process. A future analysis of the protein content of the three TET enzymes and metabolite changes (in particular α-ketoglutarate) in the two cell types could help to explain what drives the 5hmC accumulation in the differentiated villi cells.

Further validation of the RNA-seq data described in Figure 6.1 C yielded mixed results. The discrepancy observed between the RNA-seq and qRT-PCR methods can likely be explained by the differences in measuring the transcript levels utilised by the two experimental techniques. Additionally, the biological variability between the animals from which the cells were taken and technical variation involved in the isolation of the two cell populations could also account for some of the differences observed. Lastly, the limited reliability of qRT-PCR in detecting changes in gene expression that are under two-fold could also be a contributing factor (Karlen et al., 2007).

This chapter also describes the changes in *TET* transcripts and 5hmC levels during *in vitro* differentiation of hESCs to HLCs, which is a novel contribution to the understanding of the role of epigenetics during cellular differentiation. This research area has not yet been studied extensively and represents an exciting avenue to explore in terms of improving differentiation efficiency of the hESCs. It is particularly important in the process of generation of HLCs from the renewable resource of hESCs. As alluded to before, these cells hold significant promise for both liver transplantation procedures, which are the only treatment option for patients with liver failure, and toxicity assessment during drug development (Hay et al., 2008; Yu et al., 2012). As their name suggests, however, these cells are only “like” human hepatocytes and exhibit numerus

differences compared to them, most notably in the expression of the CYP enzymes. In fact, they were described to resemble the foetal hepatocytes more closely than adult ones (Baxter et al., 2015). Given that 5mC and 5hmC are known to play critical roles in regulating gene expression during differentiation, they represent not only a possible target for improving the differentiation efficiency, but also for making the HLCs resemble human adult hepatocytes more closely than they currently do.

The H9 ESCs were characterised by high *TET1* expression, with *TET2* and *TET3* genes being transcribed at much lower level and *TET3* being more abundant than *TET2*. This is in line with observations made by Li and colleagues, who differentiated H1 ESCs to pancreatic progenitor cells (pancreatic endoderm) and assessed the changes in global 5mC and 5hmC levels as well as the levels of *TET* transcripts (Li et al., 2018). They also described a progressive reduction in *TET1* expression combined with a corresponding increase in the levels of *TET2* and *TET3*, which is largely in line with the results shown in Figure 6.4, although no significant change in *TET3* expression was observed during the differentiation protocol described here. In a similar fashion, they reported little changes in global 5mC levels which supports the observations made in Figure 6.5. Lastly, Li and colleagues also described a "U-shape" distribution in 5hmC levels during the differentiation process (Li et al., 2018). In a similar fashion to the findings of this chapter, a drop in 5hmC content occurred in the initial stages of the differentiation which was followed by a gradual recovery to levels approaching those seen in the ESCs. This was the main difference in findings between their data and the results described in this chapter as 5hmC content in HBs and HLCs surpassed that of the H9 ESCs. This discrepancy in results could be due to the high levels of 5hmC observed in the adult liver (Li and Liu, 2011; Nestor et al., 2012), and the fact that Li and co-authors did not produce a terminally differentiated pancreatic cell in their protocol. A study of *in vitro* differentiation of a bipotent progenitor to hepatocytes further highlighted the importance of global 5hmC changes in the process (Rodríguez-Aguilera et al., 2020). The enrichment of 5hmC was found to be closely associated with genes involved in the hepatocyte-specific transcriptional landscape.

Following on from the findings of *TET* levels and 5hmC content fluctuations during the maturation of HLCs from H9 ESCs, it was hypothesised that one can utilise these novel results in an attempt at improving the efficiency of the differentiation protocol and match the cellular identity of the resulting HLCs more closely to the human adult

hepatocytes. This is supported by the evidence that modulating *Tet* levels, both *in vitro* and *in vivo*, affects the endodermal lineage differentiation (Koh et al., 2011; Dai et al., 2016). While the initial experiment detailed in Figure 6.6 was unsuccessful, further improvements could be made to better understand if this strategy could work and thus provide a proof-of-principle for further investigation. These were not completed due to the time constraints of the project. The first major limitation of the experiment was that the differentiation was terminated at the DE stage. This was chosen on purpose as it represents the shortest stage (taking only three days) of the protocol, thus making it ideal for an initial experiment. However, it is feasible that more noticeable improvements in differentiation efficiency would have been seen when the differentiating cells matured into HBs or HLCs. Additionally, only a small degree of *TET1* knockdown (a 22% reduction) was achieved at the endpoint of the assay. This could be improved in future experiments by transfection followed by antibiotic selection, which produces stably transfected clones that would likely exhibit significantly lower *TET1* levels. Lastly, combining the *TET1* knockdown with TET2 and TET3 overexpression could yield a better outcome by exacerbating the changes seen in *TET* transcript abundance in both directions.

In summary, the work described in this chapter furthers the understanding of the dynamics of the TET enzymes and 5hmC during the mouse intestinal differentiation and provides a novel contribution in the field of stem cell research by highlighting the previously unknown changes in *TET* transcripts and global hydroxymethylation levels during the formation of hepatocyte-like cells from human embryonic stem cells.

Chapter 7

Final discussion and future directions

The work described in this thesis furthers the understanding of the role of the TET proteins and 5hmC in cancer progression and cellular differentiation. The successful optimisation of the in-house mass spectrometry method of global 5mC and 5hmC detection in DNA samples provides a logical stepwise framework to follow by others aiming to study the changes in levels of these epigenetic modifications. In particular, the setup of this technique on a mass spectrometer less sensitive than ones previously used in the literature for similar applications means that this type of analysis should be available to a wider range of research groups in the future, especially ones with limited funding opportunities. This is further supported by a higher sensitivity of the in-house protocol compared to previously published external one used at King's College London. This experimental technique will also be critical for future experiments performed by the Murrell group, which will build on the findings described in this thesis. Importantly, a similar approach to the one described here can be used to establish a protocol for detection of global 5fC and 5caC levels in the future.

The detailed *in vitro* characterisation of the SW480 (primary) and SW620 (metastatic) CRC cell lines resulted in creation of a detailed baseline for the behaviour of these cells in culture. This provides a critical context for understanding the results described in this thesis and for future experiments performed using this *in vitro* CRC metastasis model. In particular, it will be critical to understand their behaviour in transwell invasion assays to complement the migration data reported here. This should be further followed by assessment of the *in vivo* metastatic properties of the two cell lines, either in zebrafish or mouse models. Both cell lines have already been transfected with, and are stably expressing, a red fluorescent protein and luciferase reporters by the author and are hence ready to be used for such an experiment (data not shown). The future characterisation of SW480 and SW620 cells should also focus on assessing their metabolic profiles, and in particular 2-oxoglutarate, to better understand the effects they have on the global 5hmC levels. This could be complemented by determining the post-translational modifications affecting the activity of the three TET proteins in the two cell lines.

The anti-cancer properties of vitamin C were confirmed in the SW480 and SW620 cell lines. While the ascorbic acid treatment elevated 5hmC levels in the two cell lines, its suppressive effects on proliferation and colony formation were found to be 5hmC-independent through the use of peroxide ions scavenger catalase. It was also discovered that the vitamin C treatment did not affect the expression of the three *TET* genes in the

two CRC cell lines; a novel finding which further supports its role in increasing the 5hmC levels by affecting the activity of the TET proteins. In the future, it would be particularly interesting to extend the treatment period of the two cell lines with the low vitamin C dose, possibly to multiple weeks, to study its long-term effects. Additionally, a higher concentration of ascorbic acid supplemented with catalase should be investigated as it could lead to a greater upregulation in global 5hmC levels which could potentially provide a phenotypic effect not seen with the lower doses. Lastly, the most intriguing follow-up to this study would be to expose CRC patients to similar doses of intravenous vitamin C as previously used by Linus Pauling and measure the global 5hmC changes in tumour samples following the treatment compared to biopsies taken prior to the treatment. It is tempting to speculate that a number of epigenetic changes would be observed following such treatment that have a suppressive effect on the primary tumour but could inadvertently promote CRC metastasis in the patients. All the experiments listed above should be combined with whole genome bisulfite and ox-bisulfite sequencing to complement the global 5mC and 5hmC changes with an understanding of the regions that gain hydroxymethylation. This in turn should allow one to make more accurate predictions regarding the functional significance of these local changes. Should the results of the abovementioned investigations be promising, one can envision wide adoption of vitamin C as an adjuvant to cancer therapy given its low cost and toxicity as well as widespread availability.

The results obtained with the *TET* knockdown SW480 and SW620 cells are limited by the fact that the TET and 5hmC ablation was incomplete even in the triple *TET* knockdown cell lines. Future work should determine exactly why that is the case despite the confirmation of the presence of CRISPR/Cas9- induced mutations in all knockdown cell lines by two independent experiments. Intriguingly, the reduced TET and 5hmC levels had a different effect on the migration of the primary and metastatic CRC cell lines. The *TET* TKDs in the SW480 cells resulted in no change of the *in vitro* migration rate than the one observed in the wild type cell line, while a reduction in migration was observed for the SW620 TKD cells. It will be critical to confirm this effect both in the *in vitro* transwell invasion assay and in animal experiments comparing the *in vivo* metastatic behaviour of the TKD cell lines and the wild type counterpart. If the results of these support the findings reported here, once could speculate that there is a switch in the role of the TET enzymes and 5hmC in the tumorigenesis process. In the early stages

of tumour development, the TETs could be playing a tumour suppressive role by restricting the growth of the tumour. During the metastatic cascade, however, they could be responsible for activating the expression of genes involved in cellular motility and epithelial to mesenchymal transition cycles. In that setting, they could have an oncogenic function in CRC and potentially other cancer types. These experiments should also be supplemented by studying the effects of TET overexpression in these two cell lines. Observing the opposite effects on migration due to TET overexpression in SW480 and SW620 cells would strengthen the findings described here and reduce the likelihood of them arising due to off-target effects of the CRISPR/Cas9 system. The use of overexpression vectors encoding both the full length and catalytically inactive TET enzymes will shed a light on whether the phenotypic effects observed in the two CRC cell lines are 5hmC-dependent. It is tempting to speculate that, at least in part, these will be mediated by catalytically independent function of the TET enzymes, in particular the repressive complex of TET1 and SIN3A. Lastly, based on the current findings, it is likely that the body of knowledge regarding the functions of the TET proteins that are independent of 5mC oxidation will substantially grow in the future years.

The use of loci-specific 5hmC changes as a diagnostic tool for early cancer detection will likely be widely adopted in the future. This is a cutting-edge field that is early in its development, but the preliminary evidence is certainly promising. In particular, the use of cell-free DNA offers a non-invasive way of obtaining material for 5hmC-based diagnostic purposes. It will also allow the clinicians to detect the disease early which is critical in, for example, pancreatic cancer sufferers. The vast majority of them present with late-stage disease which is no longer operable and often has already metastasised to other organs. The main challenge for this technology is the thorough validation of tumour type-specific differentially hydroxymethylated loci in cell-free DNA that will ensure high sensitivity and specificity of the test.

The findings validating the expression of the *Tet* genes in mouse small intestine add to the body of evidence supporting the role of 5hmC in the intestinal differentiation and shed light on the controversy regarding the importance of TET1 in that process. Future research should aim to resolve how 5hmC is accumulating in the differentiated villus cells compared to the stem cell crypt progenitors despite little change seen in the expression of the three *Tet* genes between these two cell types. In particular, the assessment of the TET protein levels will either validate the *Tet* transcript expression

profile in the differentiated and stem cells or provide an explanation for the 5hmC changes observed during intestinal differentiation. It is certainly possible for the TET transcript and protein levels to show pronounced differences; this was indeed observed in some of the *TET* knockdown SW480 and SW620 cell lines described here. The assessment of intracellular metabolite and iron levels in the two cell types, as well as the oxygen gradients across a section of the mouse small intestine, will provide a clear answer to the role of the three cofactors (oxygen, iron (II) and 2-oxoglutarate) in establishing the global 5hmC patterns across the crypt-villus axis.

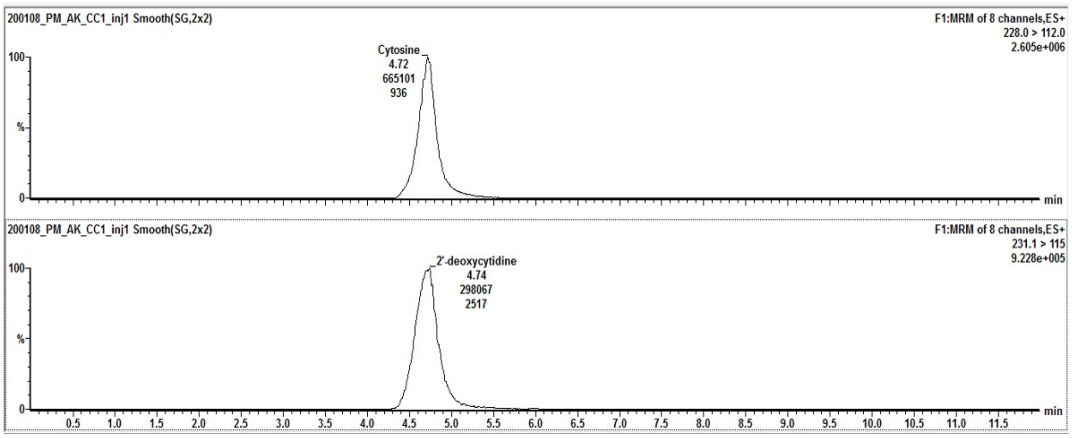
The assessment of *TET* expression and global 5mC and 5hmC changes during differentiation of hESCs to HLCs represents a novel finding that expands the growing understanding of the function of the TETs in acquisition of cellular identity. While the initial experiment aiming at utilising these results to improve hESC differentiation efficiency to definitive endoderm failed, likely due to insufficient *TET1* knockdown levels and excessive cell death associated with the transfection process, a different approach will likely yield considerably better results. The creation of a stable *TET1* knockdown hESC cell line prior to its differentiation will avoid the large amount of cell death occurring prior to the start of the differentiation process while having little effect on hESC pluripotency based on the phenotype of *TET1* knockout mESCs. This could be combined with overexpression of either TET2 or TET3 to further exacerbate the changes reported here. However, while such a crude approach could yield positive results, it is tempting to speculate that a more targeted, locus-specific induction of methylation and hydroxymethylation content changes will be used in the future to generate HLCs resembling more closely the adult hepatocytes and iPSCs that are more similar to ESCs than currently possible. This will likely be achieved through approaches such as fusion of the catalytic domain of a TET enzyme with an inactive Cas9 enzyme, which were previously successfully utilised in the literature (Choudhury et al., 2016; Xu et al., 2016; Xu et al., 2018). One can envision creating a library of guide RNA sequences that target the fusion enzyme to key loci that need to be activated in the target cell type, such as HLCs or iPSCs. If this strategy proves successful and scalable, it will significantly contribute to the fields of regenerative and personalised medicine as well as drug toxicity screening.

Almost five decades ago, the “war on cancer” was declared, and although many battles have been won, the patients still remain on the losing side. Given the focus of this thesis

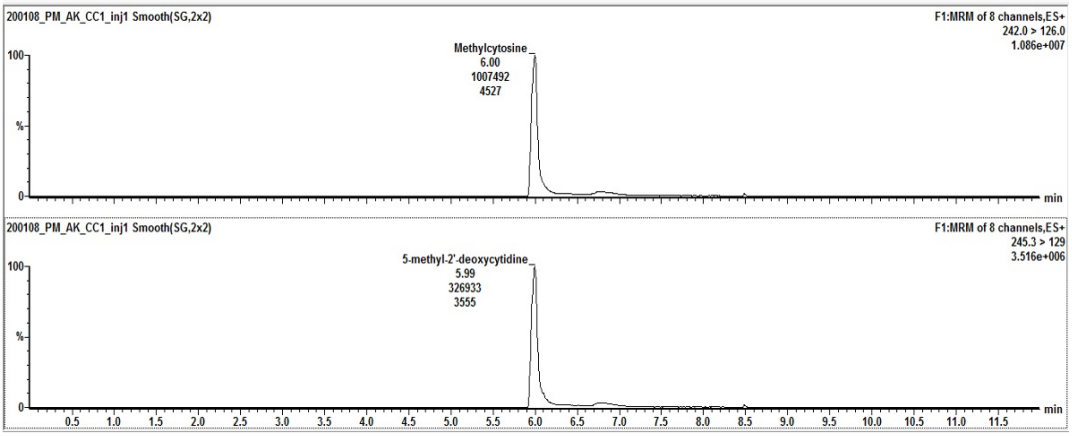
on the role of epigenetic mechanisms in CRC metastasis to the liver, it is the hope of the author that the findings described here will contribute to the field of cancer prevention. That hope reflects the simple truth that preventative measures are superior to the cure in numerous aspects, and in particular when it comes to chronic diseases such as cancer. With an increasing number of tests and treatments available to patients every year, as well as the increasing awareness regarding risk factors for tumour development, the future outlook for cancer prevention and treatment is certainly positive.

Supplementary figures

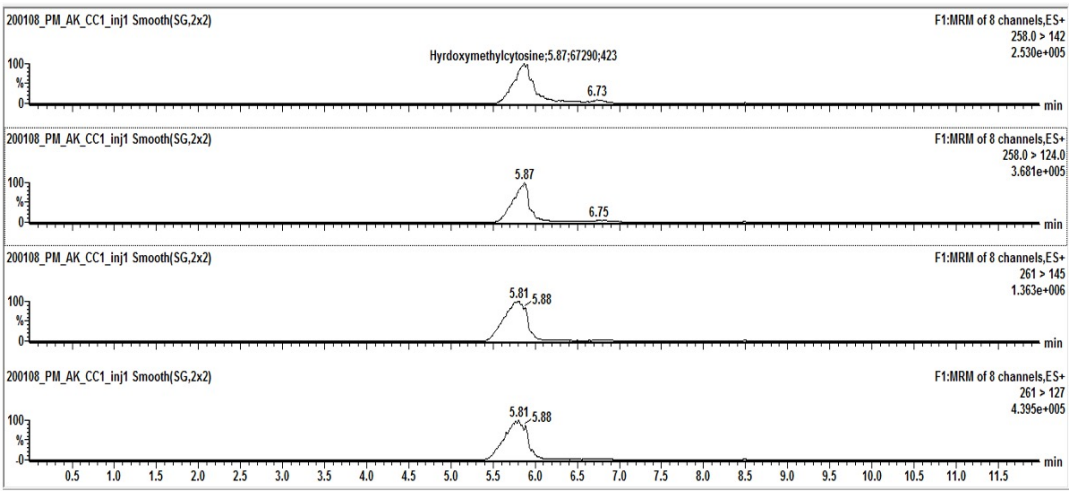
A


200108_PM_AK_CC1_inj1 Smooth(SG,2x2)
F1:MRM of 8 channels,ES+
228.0 > 112.0
2.605e+006
Cytosine
4.72
665101
936
100
%
0
min
200108_PM_AK_CC1_inj1 Smooth(SG,2x2)
F1:MRM of 8 channels,ES+
231.1 > 115
9.228e+005
2'-deoxycytidine
4.74
298067
2517
100
%
0
min
0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10.0 10.5 11.0 11.5

B


200108_PM_AK_CC1_inj1 Smooth(SG,2x2)
F1:MRM of 8 channels,ES+
242.0 > 126.0
1.086e+007
Methylcytosine
6.00
1007492
4527
100
%
0
min
200108_PM_AK_CC1_inj1 Smooth(SG,2x2)
F1:MRM of 8 channels,ES+
245.3 > 129
3.516e+006
5-methyl-2'-deoxycytidine
5.99
326933
3555
100
%
0
min
0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10.0 10.5 11.0 11.5

C



Supplementary Figure 1 The representative chromatograms of all the compounds analysed on the Xevo TqD triple quadrupole mass spectrometer

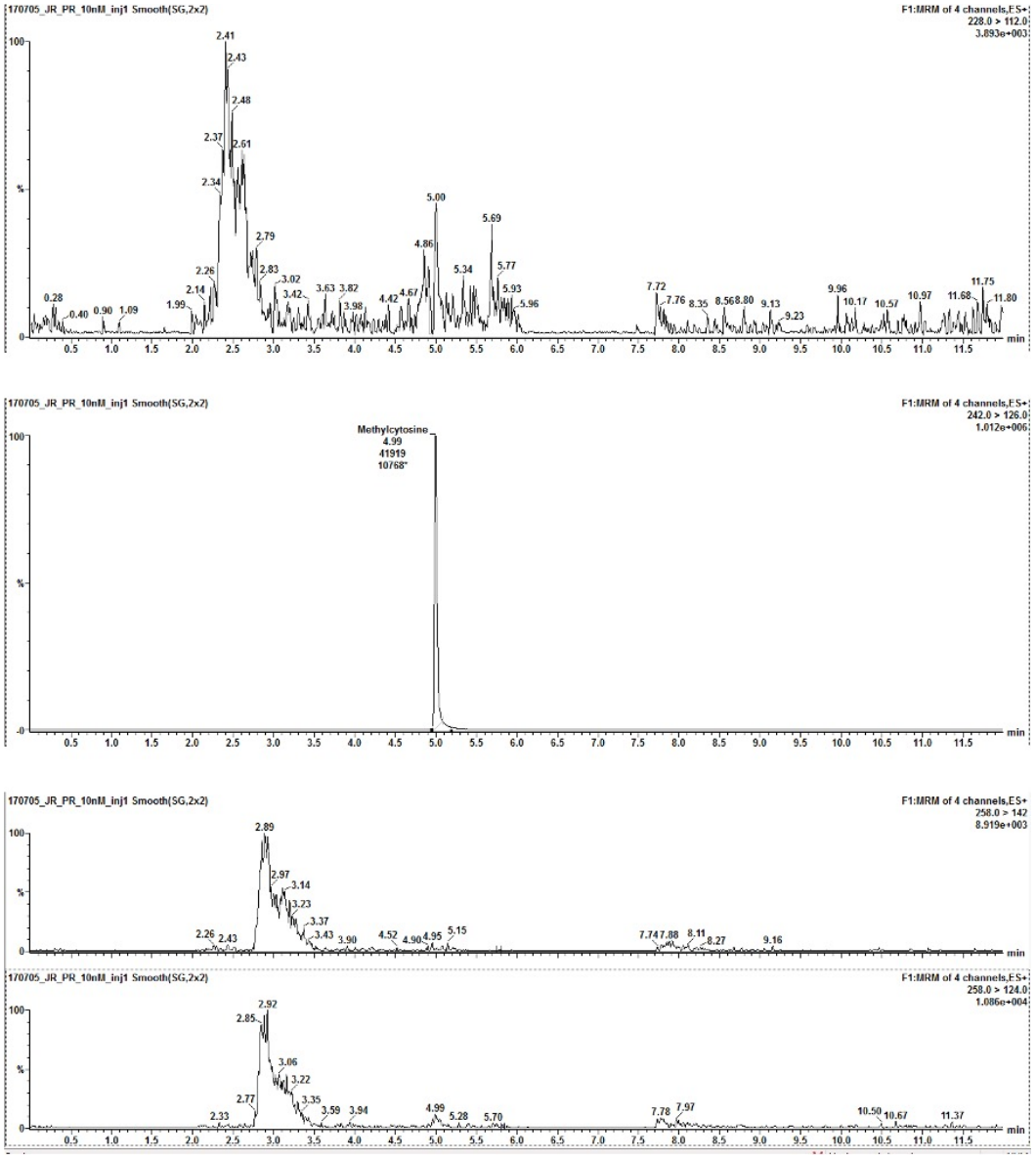
(A) The representative chromatograms for 2′-deoxycytidine (top graph) and 2'-deoxycytidine (^{15}N3) (isotopically-labelled internal standard, bottom graph).

(B) The representative chromatograms for 5-methyl-2′-deoxycytidine (top graph) and 5-methyl-2'-deoxycytidine-d3 (isotopically-labelled internal standard, bottom graph).

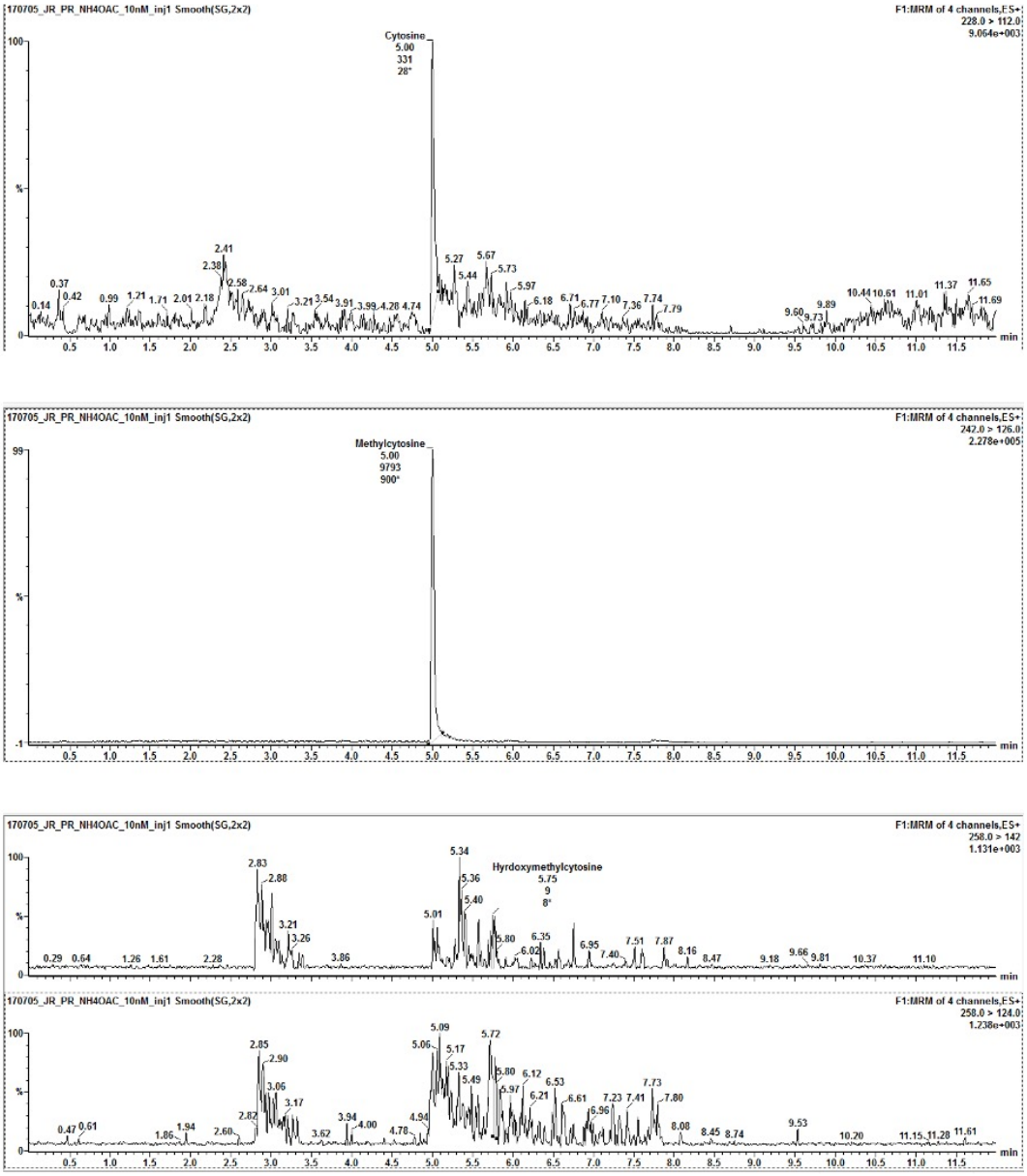
(C) The top two graphs show representative chromatograms for 5-hydroxymethyl-2′-deoxycytidine quantifier (top graph) and qualifier (bottom graph) ions. The bottom two graphs show representative chromatograms for the 5-hydroxymethyl-2'-deoxycytidine-d3 (isotopically-labelled internal standard) quantifier (top graph) and qualifier (bottom graph) ions.

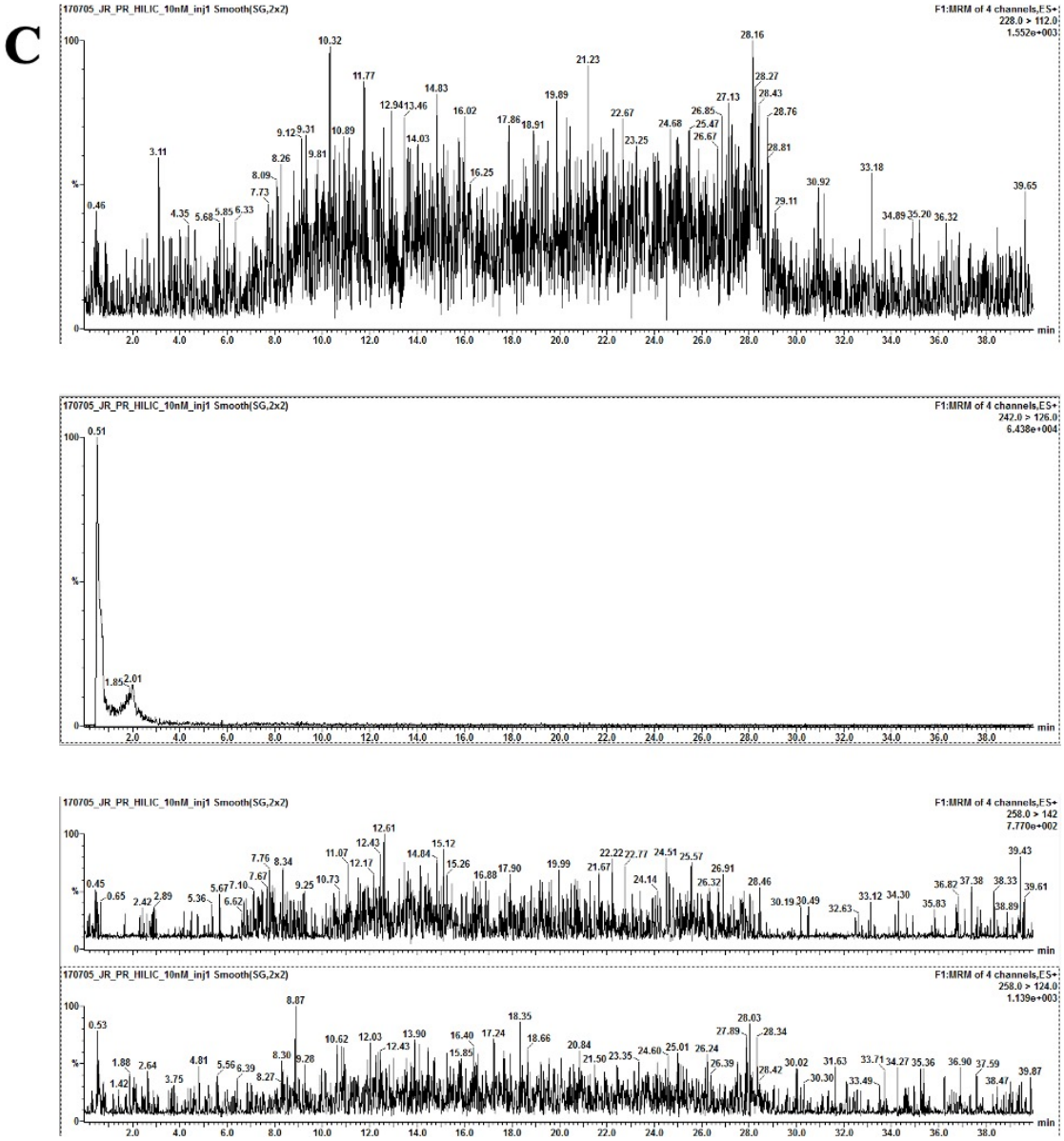
The y-axis represents the signal obtained on the mass spectrometer and the x-axis provides the retention time in minutes for all graphs.

A


170705_JR_PR_10nM_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
228.0 > 112.0
3.893e+003
Methylcytosine
4.99
41919
10768*
F1:MRM of 4 channels,ES+
242.0 > 126.0
1.012e+006
F1:MRM of 4 channels,ES+
258.0 > 142
8.919e+003
F1:MRM of 4 channels,ES+
258.0 > 124.0
1.086e+004
min

B


170705_JR_PR_NH4OAC_10nM_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
228.0 > 112.0
9.064e+003
Cytosine
5.00
331
28*
min
170705_JR_PR_NH4OAC_10nM_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
242.0 > 126.0
2.278e+005
Methylcytosine
5.00
9793
900*
min
170705_JR_PR_NH4OAC_10nM_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
258.0 > 142
1.131e+003
Hyrdoxymethylcytosine
5.75
9
8*
min
170705_JR_PR_NH4OAC_10nM_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
258.0 > 124.0
1.238e+003
min

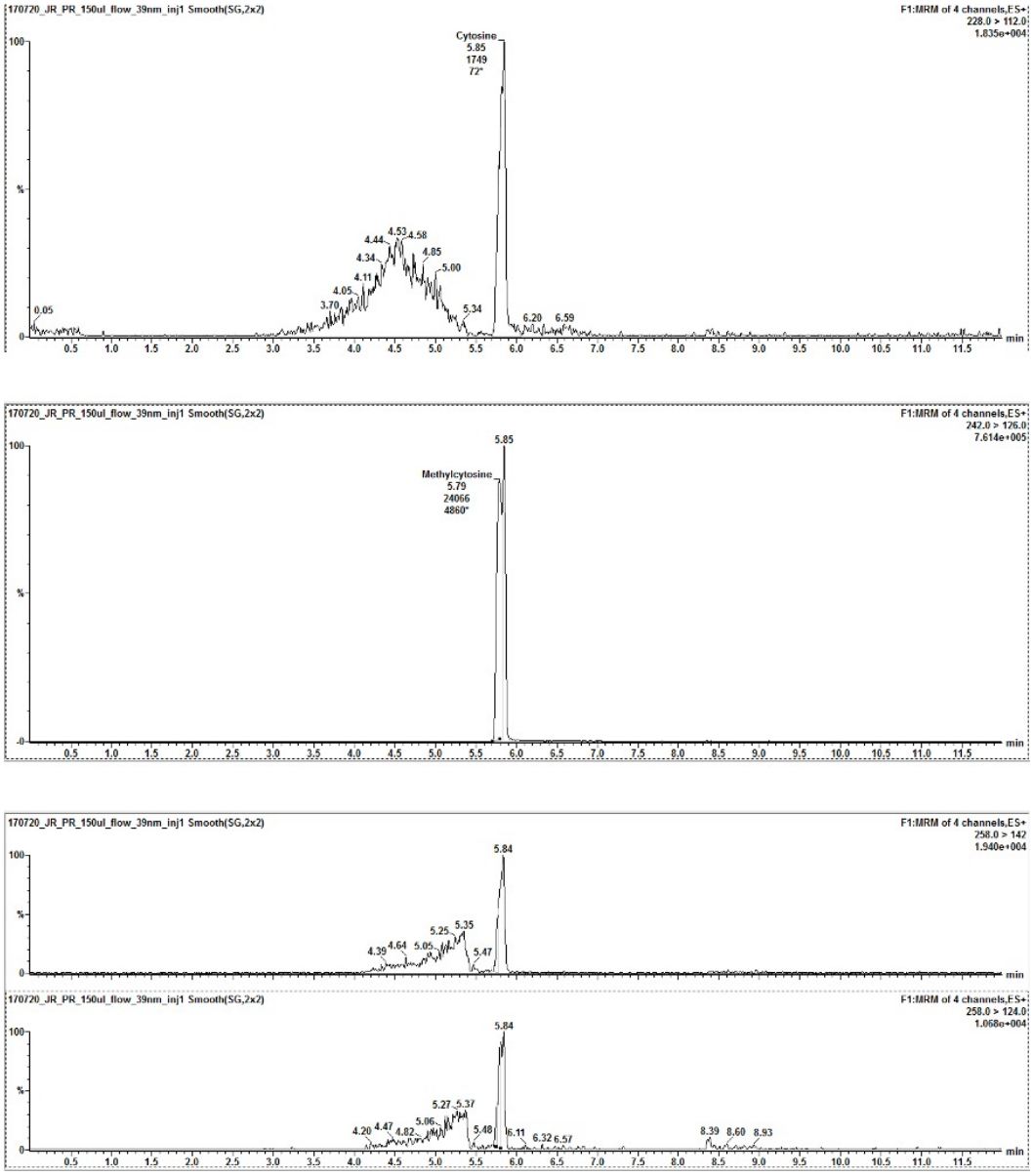


Supplementary Figure 2 The representative chromatographic results of liquid chromatography columns and conditions testing

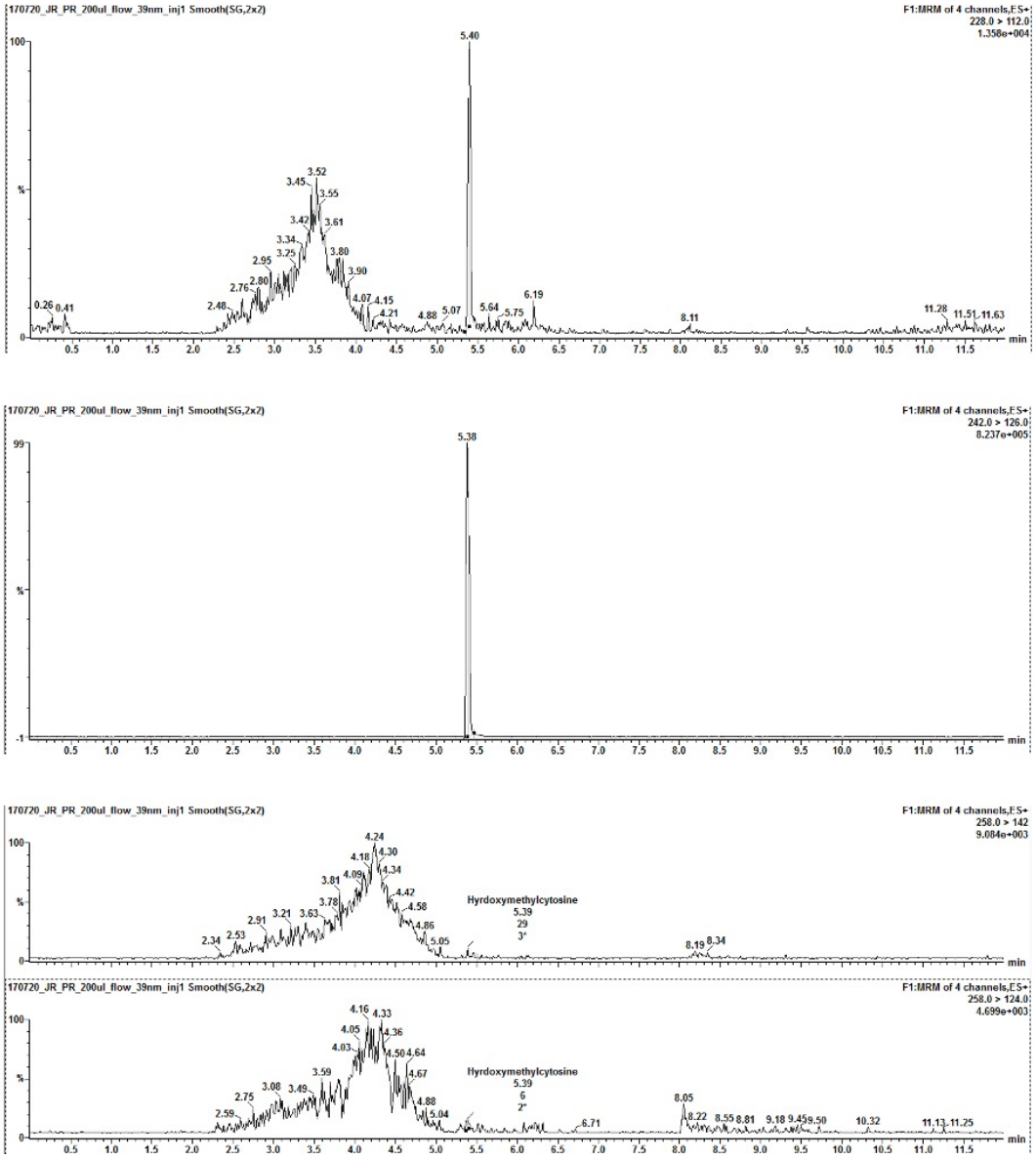
The representative chromatograms for 2′-deoxycytidine (top graph), 5-methyl-2′-deoxycytidine (middle graph) and 5-hydroxymethyl-2′-deoxycytidine quantifier and qualifier ions (bottom two graphs) for synthetic standard samples ran on: (A) the C18 column, (B) the C18 column with the addition of 10 mM ammonium acetate to the mobile phase and (C) the HILIC column.

The y-axis represents the signal obtained on the mass spectrometer and the x-axis provides the retention time in minutes for all graphs.

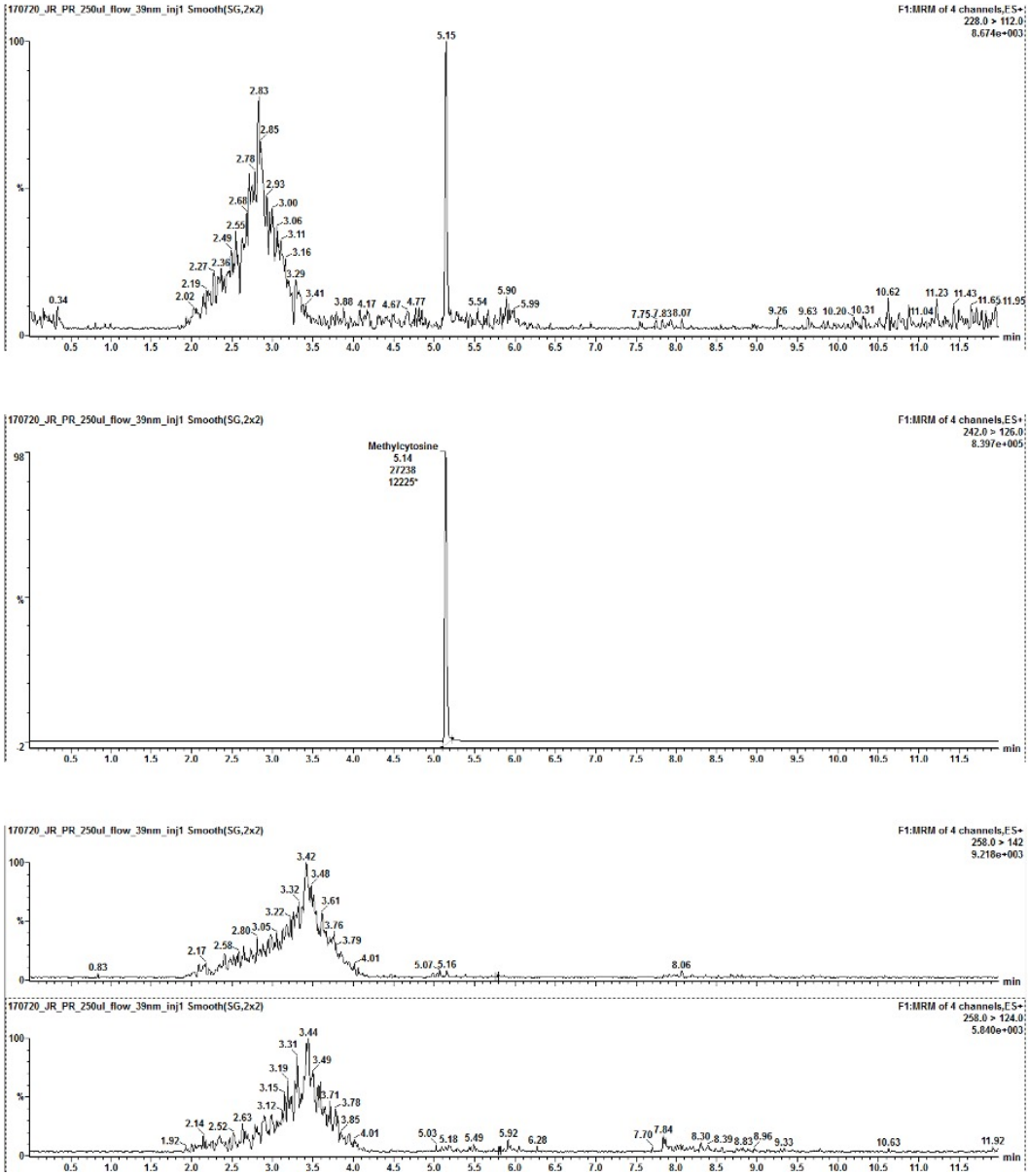
A


170720_JR_PR_150ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
228.0 > 112.0
1.835e+004
Cytosine
5.85
1749
72*
0.05
3.70
4.05
4.11
4.34
4.44
4.53
4.58
4.85
5.00
5.34
6.20
6.59
min
170720_JR_PR_150ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
242.0 > 126.0
7.614e+005
5.85
Methylcytosine
5.79
24066
4860*
min
170720_JR_PR_150ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
258.0 > 142
1.940e+004
5.84
4.39
4.64
5.05
5.25
5.35
5.47
min
170720_JR_PR_150ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
258.0 > 124.0
1.068e+004
5.84
4.20
4.47
4.82
5.06
5.27
5.37
5.48
6.11
6.32
6.57
8.39
8.60
8.93
min

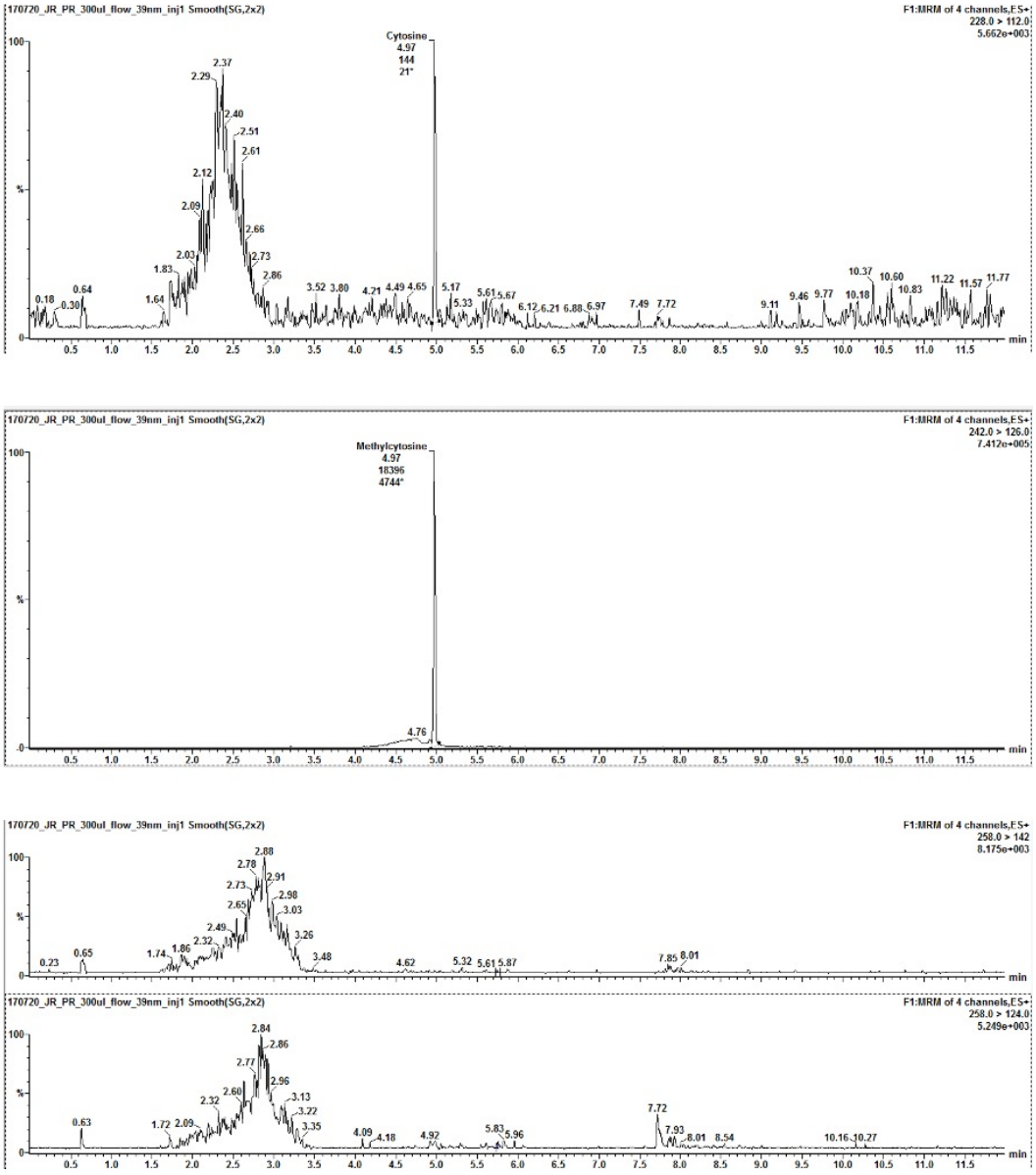
B


170720_JR_PR_200ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
228.0 > 112.0
1.358e+004
5.40
3.52
3.45
3.55
3.42
3.61
3.34
3.25
3.80
2.95
3.90
2.76
2.80
4.07
4.15
6.19
0.26
0.41
2.48
4.21
4.88
5.07
5.64
5.75
8.11
11.28
11.51
11.63
min
170720_JR_PR_200ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
242.0 > 126.0
8.237e+005
5.38
min
170720_JR_PR_200ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
258.0 > 142
9.084e+003
4.24
4.30
4.18
4.34
4.09
3.81
4.42
3.78
4.58
2.91
3.21
3.63
4.86
2.34
2.53
5.05
Hyrdoxymethylcytosine
5.39
29
3*
8.19
8.34
min
170720_JR_PR_200ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
258.0 > 124.0
4.699e+003
4.16
4.33
4.05
4.36
4.03
4.50
4.64
4.67
3.59
3.08
3.49
2.75
4.88
2.59
5.04
Hyrdoxymethylcytosine
5.39
6
2*
6.71
8.05
8.22
8.55
8.81
9.18
9.45
9.50
10.32
11.13
11.25
min

C


170720_JR_PR_250ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
228.0 > 112.0
8.674e+003
5.15
2.83
2.85
2.78
2.93
2.68
3.00
3.06
2.55
2.49
3.11
3.16
2.27
2.36
2.19
2.02
3.29
3.41
0.34
3.88
4.17
4.67
4.77
5.54
5.90
5.99
7.75
7.83
8.07
9.26
9.63
10.20
10.31
10.62
11.04
11.23
11.43
11.65
11.95
min
170720_JR_PR_250ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
242.0 > 126.0
8.397e+005
Methylcytosine
5.14
27238
12225*
min
170720_JR_PR_250ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
258.0 > 142
9.218e+003
3.42
3.48
3.32
3.61
3.22
3.76
2.80
3.05
3.79
2.17
2.58
4.01
0.83
5.07
5.16
8.06
min
170720_JR_PR_250ul_flow_39nm_inj1 Smooth(SG,2x2)
F1:MRM of 4 channels,ES+
258.0 > 124.0
5.840e+003
3.44
3.31
3.49
3.19
3.15
3.71
3.78
3.12
2.14
2.52
2.63
3.85
4.01
1.92
5.03
5.18
5.49
5.92
6.28
7.70
7.84
8.30
8.39
8.83
8.96
9.33
10.63
11.92
min

D



Supplementary Figure 3 The representative chromatographic results of liquid chromatography flow rate testing

The representative chromatograms for 2′-deoxycytidine (top graph), 5-methyl-2′-deoxycytidine (middle graph) and 5-hydroxymethyl-2′-deoxycytidine quantifier and qualifier ions (bottom two graphs) for synthetic standard samples ran through the liquid chromatography column with the following flow rates: (A) 150 µL/min, (B) 200 µL/min, (C) 250 µL/min and (D) 300 µL/min.

The y-axis represents the signal obtained on the mass spectrometer and the x-axis provides the retention time in minutes for all graphs.

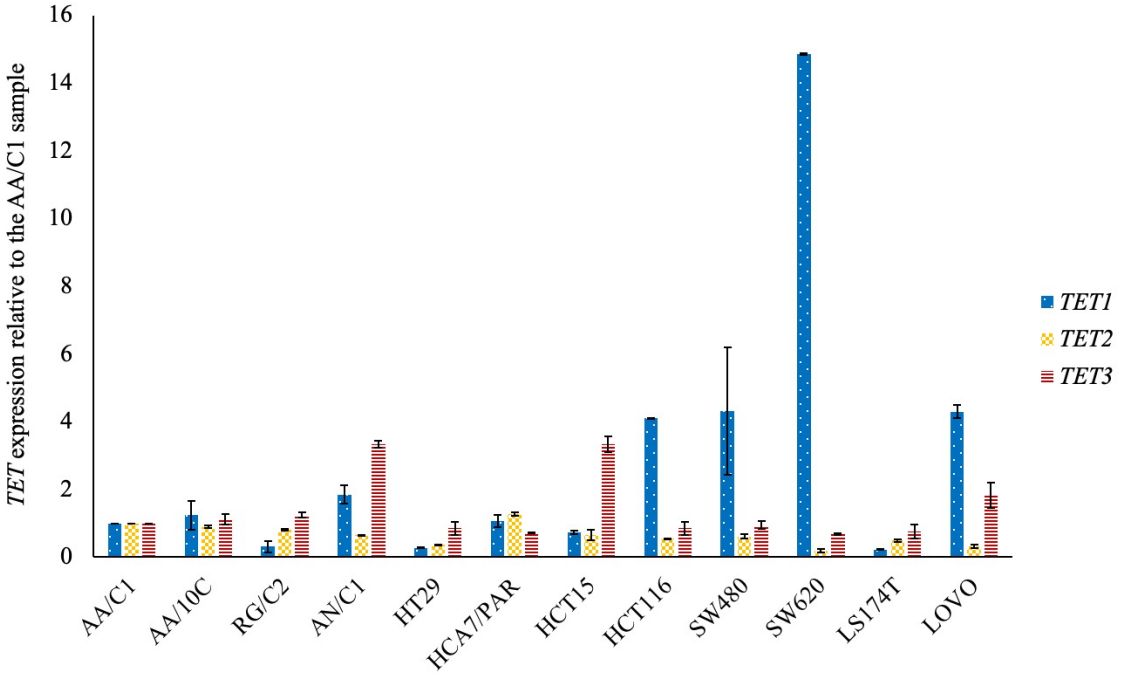
Filename	Sample Name	Cytosine Area	5mC Area	5hmC Area	Area Ratio 5mC/Cytosine	Area Ratio 5hmC/Cytosine
42314	0.5 ng/ml C, 5-mC, 5-OHm-C	288269	401474	25249		
42315	1.5 ng/ml C, 5-mC, 5-OHm-C	565345	1320616	81414		
42316	15 ng/ml C, 5-mC, 5-OHm-C	3715182	12095150	1110546		
42317	50 ng/ml C, 5-mC, 5-OHm-C	12071306	37892317	3642566		
42318	Solvent	NF	NF	NF		
42319	sample 1 HCT116 wild-type DNA1	31654261	3826534	NF	0.12089	NF
42320	sample 2 HCT116 wild-type DNA2	28748980	3406474	NF	0.11849	NF
42321	sample 3 HCT116 wild-type DNA3	24983425	3013711	NF	0.12063	NF
42322	sample 4 SW480 wild-type DNA1	38588986	4167444	NF	0.10800	NF
42323	sample 5 SW480 wild-type DNA2	35392409	3856849	NF	0.10897	NF
42324	sample 6 SW480 wild-type DNA3	32828156	3468939	NF	0.10567	NF
42325	sample 7 SW620 wild-type DNA1	39351705	3803354	NF	0.09665	NF
42326	sample 8 SW620 wild-type DNA2	55571970	5125037	NF	0.09222	NF
42327	sample 9 SW620 wild-type DNA3	41230772	3836331	NF	0.09305	NF

Supplementary Figure 4 The result summary of the external analysis of global 5-methylcytosine and 5-hydroxymethylcytosine levels in HCT116, SW480 and SW620 cell lines obtained from King's College London

The filenames 42314 - 42317 represent the synthetic standard calibration curve for the three compounds. C = cytosine, 5-mC = 5-methylcytosine, 5-Ohm-C = 5-hydroxymenthylcytosine.

The filename 42318 represents the solvent-only control containing the digestion enzyme matrix.

The filenames 42319 - 42327 describe the results of global 5mC and 5hmC levels analysis in three independently isolated DNA samples of the three colorectal cancer cell lines: HCT116, SW480 and SW620. NF = not found.



Supplementary Figure 5 The expression levels of the three *TET* genes in a panel of twelve colorectal cancer cell lines

The *TET* expression was normalized to the *TBP* reference gene. The graph represents the results of the analysis of one biological replicate for each cell line. The error bars represent the standard deviation of two technical replicates for each sample.

BACKGROUND

TET1 or tet methylcytosine dioxygenase 1 (ENSG00000138336) is located on chromosome 10: 68,560,656-68,694,482 on the forward strand. One protein coding transcript that is supported by CCDS is annotated for TET1, TET1-201 (ENST00000373644.4, CCDS7281). This transcript codes for a 2136 amino acid protein. Two paralogues are also listed which show up to 25% identity and a duplicated pseudogene (up to 66% identity) is also documented (Ensembl release 90 - August 2017). Data from Hart T *et al.,* Cell. 2015 Dec 3;163(6):1515-26 suggests that TET1 may be essential in some cell lines. However, mice with loss of TET1 are viable (Dawlaty MM, *et al.,* Cell Stem Cell. 2011 Aug 5;9(2):166-75).

TET2 or tet methylcytosine dioxygenase 2 (ENSG00000168769) is located on chromosome 4: 105,145,875-105,279,816 on the forward strand. Of the 6 protein coding transcripts listed on ensemble, 3 are supported to CCDS annotation; TET2-209 (ENST00000540549.5; CCDS47120), TET2-203 (ENST00000380013.8, CCDS47120) and TET2-202 (ENST00000305737.6, CCDS3666). TET2-209 and TET2-203 code for a 2002 amino acid protein which is considered the canonical isoform. Two paralogues of less than 30% identity are described, but no pseudogenes are documented. (Ensembl release 90 - August 2017). Data from is not essential in the cell lines examined. However, mice with loss of TET2 have been described Hart T *et al.,* Cell. 2015 Dec 3;163(6):1515-26 suggests that TET2 (Li Z, *et al.,* Blood. 2011 Oct 27;118(17):4509-18).

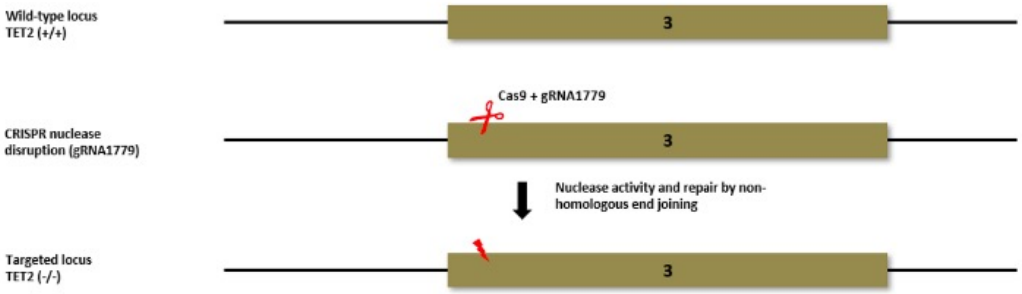
TET3 or tet methylcytosine dioxygenase 3 (ENSG00000187605) is located on chromosome 2: 73,986,404-74,108,176 on the forward strand. A single protein coding transcript TET3-202 (ENST00000409262.7) is supported by CCDS annotation (CCDS46339). This transcript codes for a 1795 amino acid protein. As for the other TET genes two paralogue are described (Ensembl release 90 - August 2017). Data from Hart T *et al.,* Cell. 2015 Dec 3;163(6):1515-26 suggests that TET3 is not essential in the cell lines examined. However, mice with partial loss of TET3 have been described but female animals are less fertile (Gu TP, *et al.,* Nature. 2011 Sep 29;477(7366):606-10).

OVERVIEW

Cell line generation was performed by targeting the loci of interest via non-homologous end joining repair of Cas9 nuclease-mediated double strand DNA breaks in three sequential targeting rounds. In the first round, the TET2 locus was targeted. One targeted clone with out-of-frame modifications to all alleles (6F9) was selected to undergo the second targeting event. In the second round, the TET3 locus was targeted. One targeted clone with out-of-frame modifications to all alleles (22G14) was selected to undergo the third targeting event. In the third round, the TET1 locus was targeted. Genetic validation was completed to identify clones with the desired genotype. The results of such genetic validation are detailed in the following data-pack.

TARGETING STRATEGY – TET2 LOCUS

Transcript to be targeted: ENST00000305737.6 (Ensembl Release 90: August 2017)
Cell line to be targeted: SW480 (HD PAR-1705, provided by the client)
Desired cell line genotype: Homozygous knock-out of TET2; TET2 (-/-)



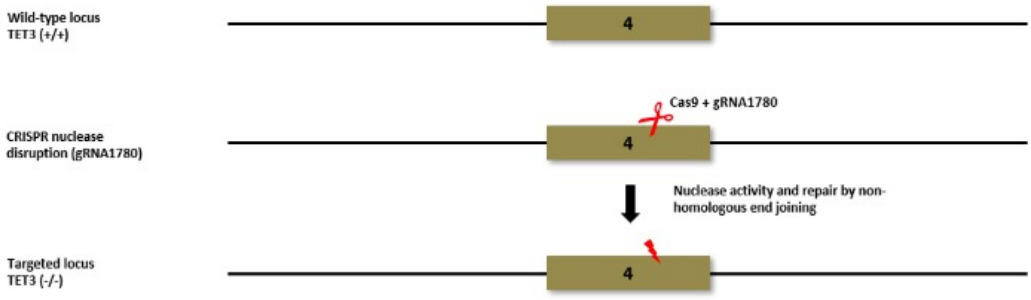
TET2 gDNA CCCTGTATGAAGGGAAGCCAGAATAGTCGTGTGAGTCCTGACTTTACACAAGA

gRNA1779 binds to green site in sequence. Blue indicates the PAM site which is also required for cutting activity. The mechanism of repair of double strand breaks will result in insertions or deletions at the gRNA cut site. If such indels cause a frame-shift, this will likely result in an early STOP codon and non-expression of the gene at the protein level.

gRNA1779 binding site
PAM site

TARGETING STRATEGY – TET3 LOCUS

Transcript to be targeted: ENST00000409262.7 (Ensembl Release 90: August 2017)
Cell line to be targeted: SW480 TET2 (-/-); HD 283-001 clone 6F9
Desired cell line genotype: Homozygous knock-out of TET2 and TET3; TET2 (-/-) TET3 (-/-/-)



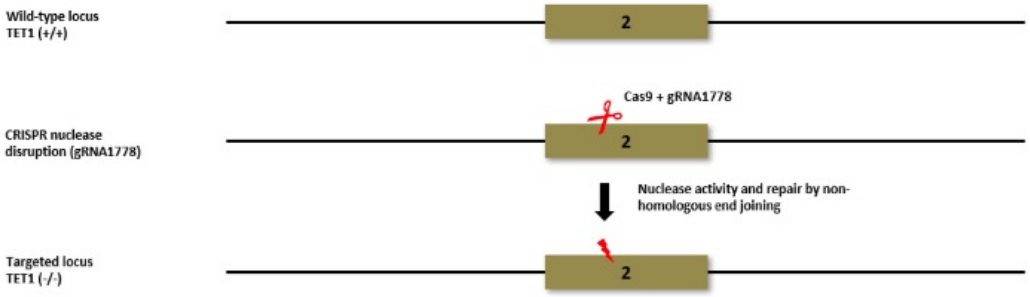
TET3 gDNA CTCACTTGGGATCTGGCCCCACGGTCGCCTCTATCCGGGAACTCATGGAGGAG

gRNA1780 binds to green site in sequence. Blue indicates the PAM site which is also required for cutting activity. The mechanism of repair of double strand breaks will result in insertions or deletions at the gRNA cut site. If such indels cause a frame-shift, this will likely result in an early STOP codon and non-expression of the gene at the protein level.

gRNA1780 binding site
PAM site

TARGETING STRATEGY – TET1 LOCUS

Transcript to be targeted: ENST00000373644.4 (Ensembl Release 90: August 2017)
Cell line to be targeted: SW480 TET2 (-/-) TET3 (-/-/-); HD 283-002 clone 22G14
Desired cell line genotype: Homozygous knock-out of TET1, TET2 and TET3; TET1 (-/-) TET2 (-/-) TET3 (-/-/-)



TET1 gDNA GGCGACTCTCCCAACCCCCACTGGTCGTAGCCAAATCCAAAAAGGTTCCACTT

gRNA1778 binds to green site in sequence. Blue indicates the PAM site which is also required for cutting activity. The mechanism of repair of double strand breaks will result in insertions or deletions at the gRNA cut site. If such indels cause a frame-shift, this will likely result in an early STOP codon and non-expression of the gene at the protein level.

gRNA1778 binding site
PAM site

CELL LINE DETAILS

Growth Conditions

Cell line	SW480
Source	University of Bath
Description	Human colorectal adenocarcinoma cell line
Media	DMEM, high glucose, GlutaMAX™ Supplement (ThermoFisher 61965026) supplemented with 10% fetal bovine serum

Clone Details

Clone Name	Catalogue Number	HD Clone Number	SNB	Genotype
SW480 Parental	HD PAR-1705	27065	29033	TET1 (+/+) TET2 (+/+) TET3 (+/+/+)
6F9	HD 283-001	29498	30732	TET1 (+/+) TET2 (-/-) TET3 (+/+/+)
22G14	HD 283-002	29646	31460	TET1 (+/+) TET2 (-/-) TET3 (-/-/-)
2I7	HD 283-003	29929	32425	TET1 (-/-) TET2 (-/-) TET3 (-/-/-)
6B12	HD 283-003	29930	32426	TET1 (-/-) TET2 (-/-) TET3 (-/-/-)

TARGETING STATISTICS

Gene	# of Colonies Screened	# of +ve Wells Identified	Assay Configuration	Non Homologous End Joining Frequency†	Results
TET2	430	66	Consolidation of populations from 10 x 384 well plates	15.3 %	Of the 66 putative targeted clones identified in the initial screens, 14 clones were taken forward for further genetic validation. 5 of these were confirmed as having out of frame modifications in both alleles of TET2. One clone (6F9) was selected for banking and validation based on further confirmatory genotyping. This clone was used for targeting at the TET3 locus.
TET3	384	75	Consolidation of populations from 30 x 384 well plates	19.5%	Of the 75 putative targeted clones identified in the initial screen, 48 were taken forward for further genetic validation. 13 of these were confirmed as having out of frame modifications in all alleles of TET3. One clone (22G14) was selected for banking and validation based on further confirmatory genotyping. This clone was used for targeting at the TET1 locus.
TET1	368	73	Consolidation of populations from 15 x 384 well plates	19.8%	Of the 73 putative targeted clones identified in the initial screen, 25 clones were taken forward for further genetic validation. 8 of these were confirmed as having out of frame modifications in both alleles of TET1. Two clones (2I7 and 6B12) were selected for banking and validation based on further confirmatory genotyping.

† Where non homologous end joining frequency is the percentage of clones identified to have had at least one allele of the gene of interest disrupted by CRISPR

GENOTYPING SUMMARY

Clone Name	Genotype	Allele	gDNA modification	Predicted impact on protein
6F9	TET2 (-/-)	1	10 bp deletion	Early STOP at position 91
		2	19 bp deletion	Early STOP at position 88
22G14	TET2 (-/-)	1	10 bp deletion	Early STOP at position 91
		2	19 bp deletion	Early STOP at position 88
	TET3 (-/-/-)	1	2 bp insertion	Early STOP at position 852
		2	10 bp deletion + 2 bp substitution (TG>AT)	Early STOP at position 891
		3	28 bp deletion	Early STOP at position 843
2I7	TET1 (-/-)	1	13 bp deletion	Early STOP at position 130
		2	16 bp deletion	Early STOP at position 129
	TET2 (-/-)	1	10 bp deletion	Early STOP at position 91
		2	19 bp deletion	Early STOP at position 88
	TET3 (-/-/-)	1	2 bp insertion	Early STOP at position 852
		2	10 bp deletion + 2 bp substitution (TG>AT)	Early STOP at position 891
		3	28 bp deletion	Early STOP at position 843
6B12	TET1 (-/-)	1	1 bp insertion	Early STOP at position 132
		2	4 bp deletion	Early STOP at position 121
	TET2 (-/-)	1	10 bp deletion	Early STOP at position 91
		2	19 bp deletion	Early STOP at position 88
	TET3 (-/-/-)	1	2 bp insertion	Early STOP at position 852
		2	10 bp deletion + 2 bp substitution (TG>AT)	Early STOP at position 891
		3	28 bp deletion	Early STOP at position 843

MYCOPLASMA SCREENING RESULTS

Clone Name	Catalogue Number	HD Clone Number	SNB	Mycoplasma (Hoechst test)
SW480 Parental	HD PAR-1705	27065	29033	Passed
6F9	HD 283-001	29498	30732	Passed
22G14	HD 283-002	29646	31460	Passed
2I7	HD 283-003	29929	32425	Passed
6B12	HD 283-003	29930	32426	Passed

CELL LINE DETAILS

Growth Conditions

Cell line	SW620
Source	University of Bath
Description	Human colorectal adenocarcinoma cell line
Media	DMEM, high glucose, GlutaMAX™ Supplement (ThermoFisher 61965026) supplemented with 10% fetal bovine serum

Clone Details

Clone Name	Catalogue Number	HD Clone Number	SNB	Genotype
SW620 Parental	HD PAR-1706	27066	29035	TET1 (+/+) TET2 (+) TET3 (+/+)
11D14	HD 282-001	29450	30572	TET1 (+/+) TET2 (-) TET3 (+/+)
1G13	HD 282-002	29696	31573	TET1 (+/+) TET2 (-) TET3 (-/-)
2H13	HD 282-003	29743	31794	TET1 (-/-) TET2 (-) TET3 (-/-)

TARGETING STATISTICS

Gene	# of Colonies Screened	# of +ve Wells Identified	Assay Configuration	Non Homologous End Joining Frequency†	Results
TET2	96	6	Consolidation of populations from 10 x 384 well plates	6.3%	Of the 6 putative targeted clones identified in the initial screen, all 6 clones were taken forward for further genetic validation. 3 of these were confirmed as having out of frame modifications in both alleles of TET2. One clone (11D14) was selected for banking and validation based on further confirmatory genotyping. This clone was used for targeting at the TET3 locus.
TET3	384	68	Consolidation of populations from 15 x 384 well plates	17.7%	Of the 68 putative targeted clones identified in the initial screen, 28 were taken forward for further genetic validation. 5 of these were confirmed as having out of frame modifications in all alleles of TET3. One clone (1G13) was selected for banking and validation based on further confirmatory genotyping. This clone was used for targeting at the TET1 locus.
TET1	384	37	Consolidation of populations from 10 x 384 well plates	9.6%	Of the 37 putative targeted clones identified in the initial screen, 7 clones were taken forward for further genetic validation. Initial genotyping indicated that 3 of these clones contained out of frame modifications in both alleles of TET1. Two clones (2H13 and 3E17) were selected for banking and validation based on further confirmatory genotyping.

† Where non homologous end joining frequency is the percentage of clones identified to have had at least one allele of the gene of interest disrupted by CRISPR

GENOTYPING SUMMARY

Clone Name	Genotype	Allele	gDNA modification	Predicted impact on protein
11D14	TET2 (-)	1	16 bp deletion	Early STOP at position 69
1G13	TET2 (-)	1	16 bp deletion	Early STOP at position 69
	TET3 (-/-)	1	1 bp insertion	Early STOP at position 934
		2	11 bp deletion	Early STOP at position 930
2H13	TET1 (-/-)	1	33 bp deletion	11 amino acid deletion
		2	12 bp (overall) insertion	5 amino acid insertion
	TET2 (-)	1	16 bp deletion	Early STOP at position 69
	TET3 (-/-)	1	1 bp insertion	Early STOP at position 934
		2	11 bp deletion	Early STOP at position 930

MYCOPLASMA SCREENING RESULTS

Clone Name	Catalogue Number	HD Clone Number	SNB	Mycoplasma (Hoechst test)
SW620 Parental	HD PAR-1706	27066	29035	Passed
11D14	HD 282-001	29450	30572	Passed
1G13	HD 282-002	29696	31573	Passed
2H13	HD 282-003	29743	31794	Passed

SUMMARY

- Non homologous end joining frequency of 6.3% was observed for TET2, 17.7% for TET3 and 9.6% for TET1.
- Clone 11D14 likely carries one TET2 allele with a 16 bp deletion in exon 3 of transcript ENST00000305737.6. Such modifications are out-of-frame changes.
- Clone 1G13 likely carries one TET2 allele with a 16 bp deletion in exon 3 of transcript ENST00000305737.6. Clone 1G13 likely carries one TET3 allele with a 1 bp insertion and a second allele with a 11 bp deletion in exon 4 of transcript ENST00000409262.7. Such modifications are out-of-frame changes.
- Clone 2H13 likely carries one TET2 allele with a 16 bp deletion in exon 3 of transcript ENST00000305737.6. Clone 2H13 likely carries one TET3 allele with a 1 bp insertion and a second allele with a 11 bp deletion in exon 4 of transcript ENST00000409262.7. Such modifications are out-of-frame changes. Clone 2H13 likely carries one TET1 allele with a 33 bp deletion and a second allele with a 12 bp insertion in exon 2 of transcript ENST00000373644.4. Such modifications are in-frame changes.
- All data generated supports the conclusion that TET2, TET3 homozygous knock-outs are present in clone 2H13 however TET1 carries an in frame modification which cannot be conclusively confirmed to result in a functional knock out of TET1.

Supplementary Figure 6 The details of the strategy used by Horizon Discovery to generate the *TET* knockdown SW480 and SW620 cell lines

The details presented in the figure were obtained from Horizon Discovery.

A

Sample	Sequencing result
SW480 WT	CCCAACCCCCACTGGTCGTAGCCAAATCCAAAAAGGTTCCACTTT
SW480 SKD	CCCAACCCCCACTGGTCGTAGCCAAATCCAAAAAGGTTCCACTTT
SW480 DKD	CCCAACCCCCACTGGTCGTAGCCAAATCCAAAAAGGTTCCACTTT
SW480 TKD 1*	CCCAACCCCC----------TCCTCCAANGTTCTTCTTN
SW480 TKD 2	CCCAACCCCCACTG-ACGTANNCAANNCCANAAAGGTTCCACTTT
SW620 WT	CCCAACCCCCACTGGTCGTAGCCAAATCCAAAAAGGTTCCACTTT
SW620 SKD	CCCAACCCCCACTGGTCGTAGCCAAATCCAAAAAGGTTCCACTTT
SW620 DKD	CCCAACCCCCACTGGTCGTAGCCAAATCCAAAAAGGTTCCACTTT
SW620 TKD*	CGTTCCCCTTTCTT----------CNGGTTTATAAAGGAACATGATT

B

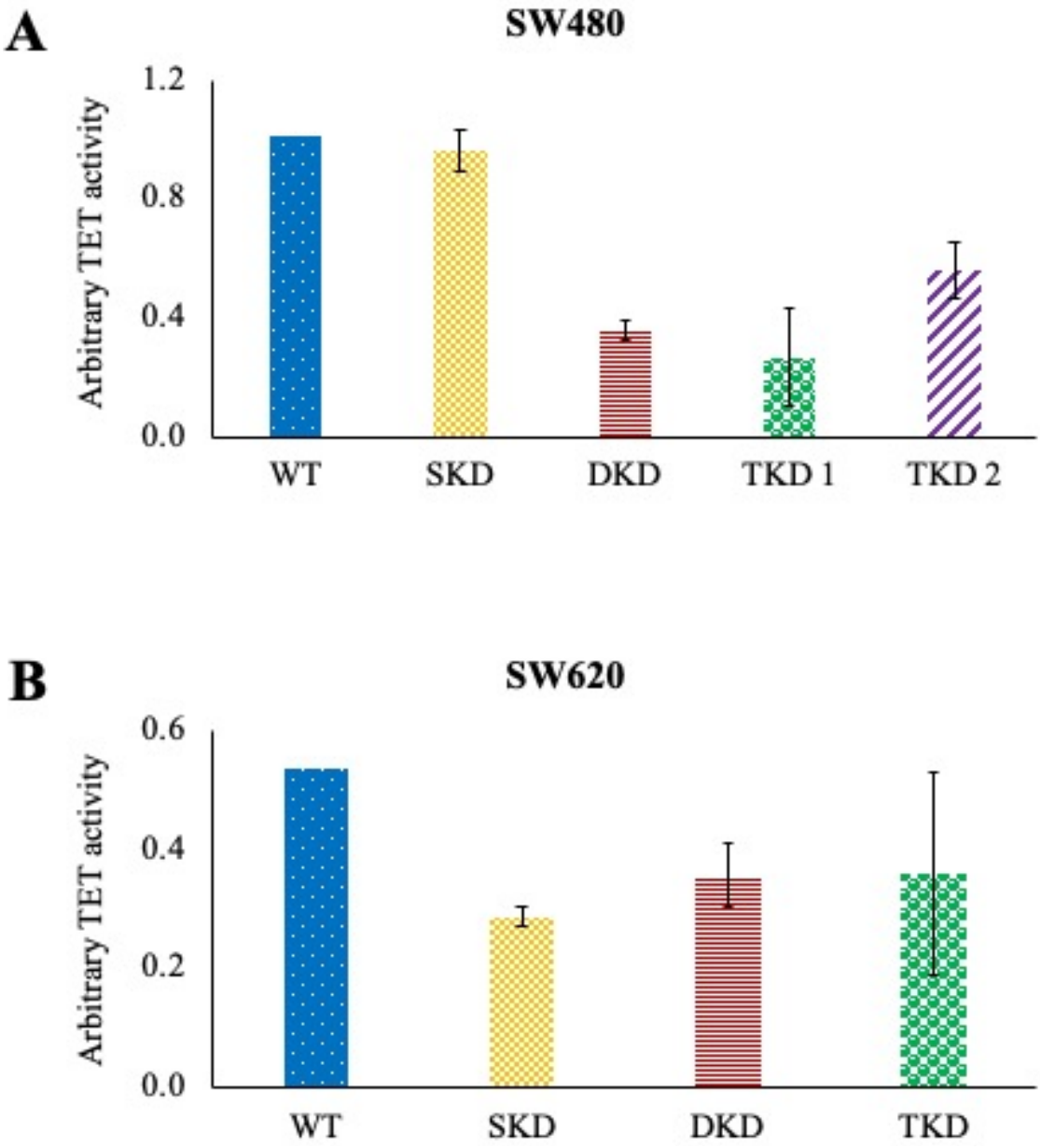
Sample	Sequencing result
SW480 WT	TGAAGGGAAGCCAGAATAGTCGTGTGAGTCCTGACTTTACACAAGAAA
SW480 SKD	TGAAGGGAAGCC----------TGACNNTNCNNACTATA
SW480 DKD	TTCCGCGTTAGCA----------ACTTTNCCTCNNNNN
SW480 TKD 1	TGAAGGGAAGCC----------TGACTNTACNNACTATA
SW480 TKD 2	TGAAGGGAAGCC----------TGACTNTACNGACNTAA
SW620 WT	TGAAGGGAAGCCAGAATAGTCGTGTGAGTCCTGACTTTACACAAGAAA
SW620 SKD**	TGAACGNNNGNNAGAAGAGTCGTGTGAGTCCTGACTTTANNNNNNAAA
SW620 DKD	T----------AGTCGTGTGAGTCCTGACTTTACACAAGAAA
SW620 TKD	T----------AGTCGTGTGAGTCCTGACTTTACACANNNNN

C

Sample	Sequencing result
SW480 WT	GGGATCTGGCCCCACGGTCGCCTCTATCCGGGAACTCATGGAGGA
SW480 SKD	GGGATCTGGCCCCACGGTCGCCTCTATCCGGGAACTCATGGAGGA
SW480 DKD*	GGGATC----------NNNNNNNNNNNNCGGTGNNNNATGCAGGA
SW480 TKD 1*	GGGATN----------NNGANNNNNNNNCNNNNNNNGATGCAGGA
SW480 TKD 2*	GGGATC----------CNGACNNNNNNNCNNNGANNGATGCANNA
SW620 WT	GGGATCTGGCCCCACGGTCGCCTCTATCCGGGAACTCATGGAGGA
SW620 SKD	GGGATCTGGCCCCACGGTCGCCTCTATCCGGGAACTCATGGAGGA
SW620 DKD	GGGATCTGGC----------CTCTATNNNNGANNNNATGCAGGA
SW620 TKD	GGGATCTGGC----------CTCTATNNNCGNNNNATGCAGGA

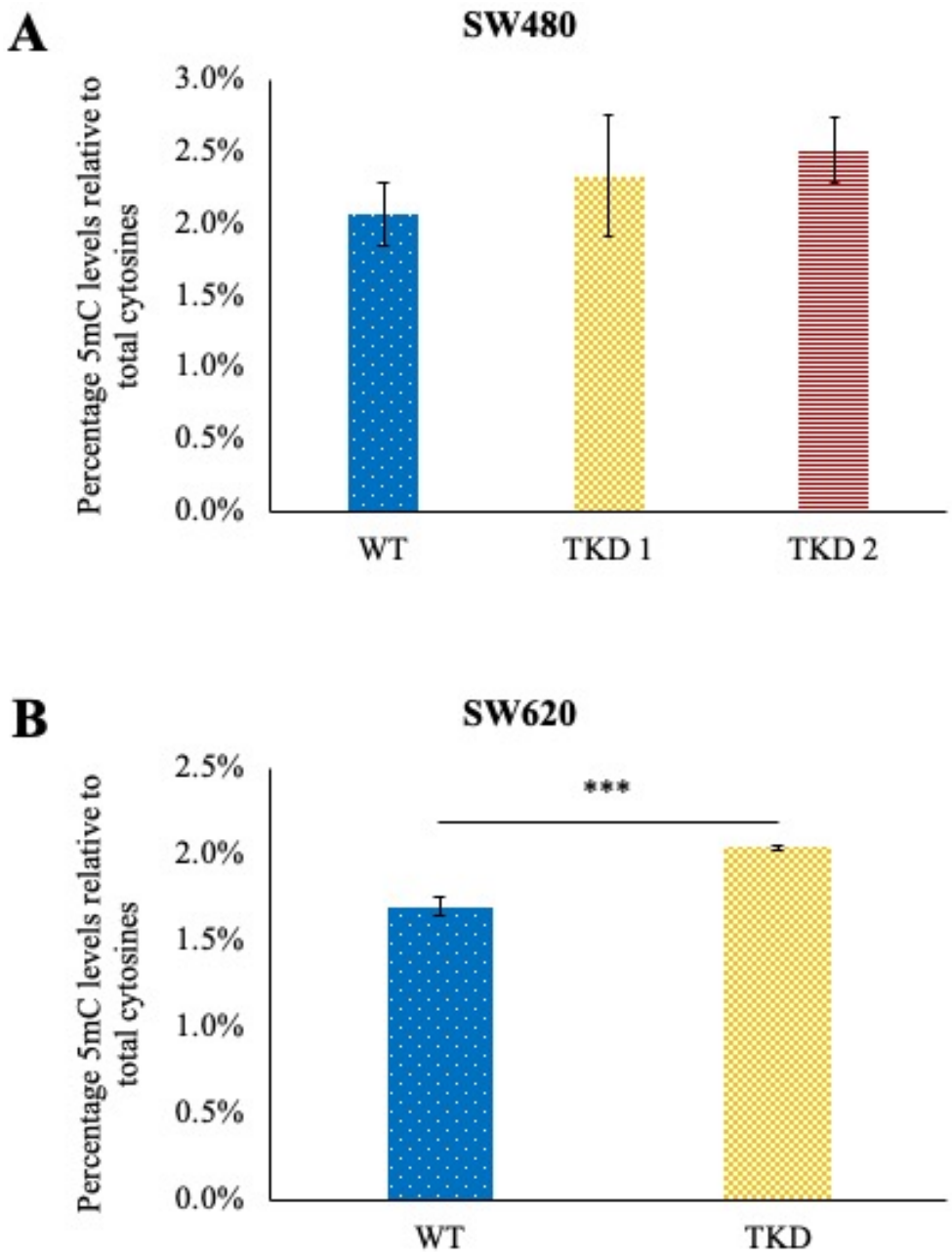
Supplementary Figure 7 The sequencing results of the *TET* knockdown SW480 and SW620 cell lines

The results of sequencing of PCR products containing the regions targeted for CRISPR/Cas9-mediated deletions of (A) *TET1*, (B) *TET2* and (C) *TET3* in wild type (WT), single (SKD), double (DKD) and triple (TKD) *TET* knockdown SW480 and SW620 cell lines. The PAM and guide RNA sequences are highlighted in blue and yellow, respectively. The muted regions are denoted by the hyphens highlighted in red. * = the samples showing extensive scarring around the deletion site. ** = the sample does not show any deletions despite harbouring the *TET2* mutation as shown by the sequencing of SW620 DKD and TKD samples; this is likely an experimental error or the wild type allele was sequenced instead of the mutated one.



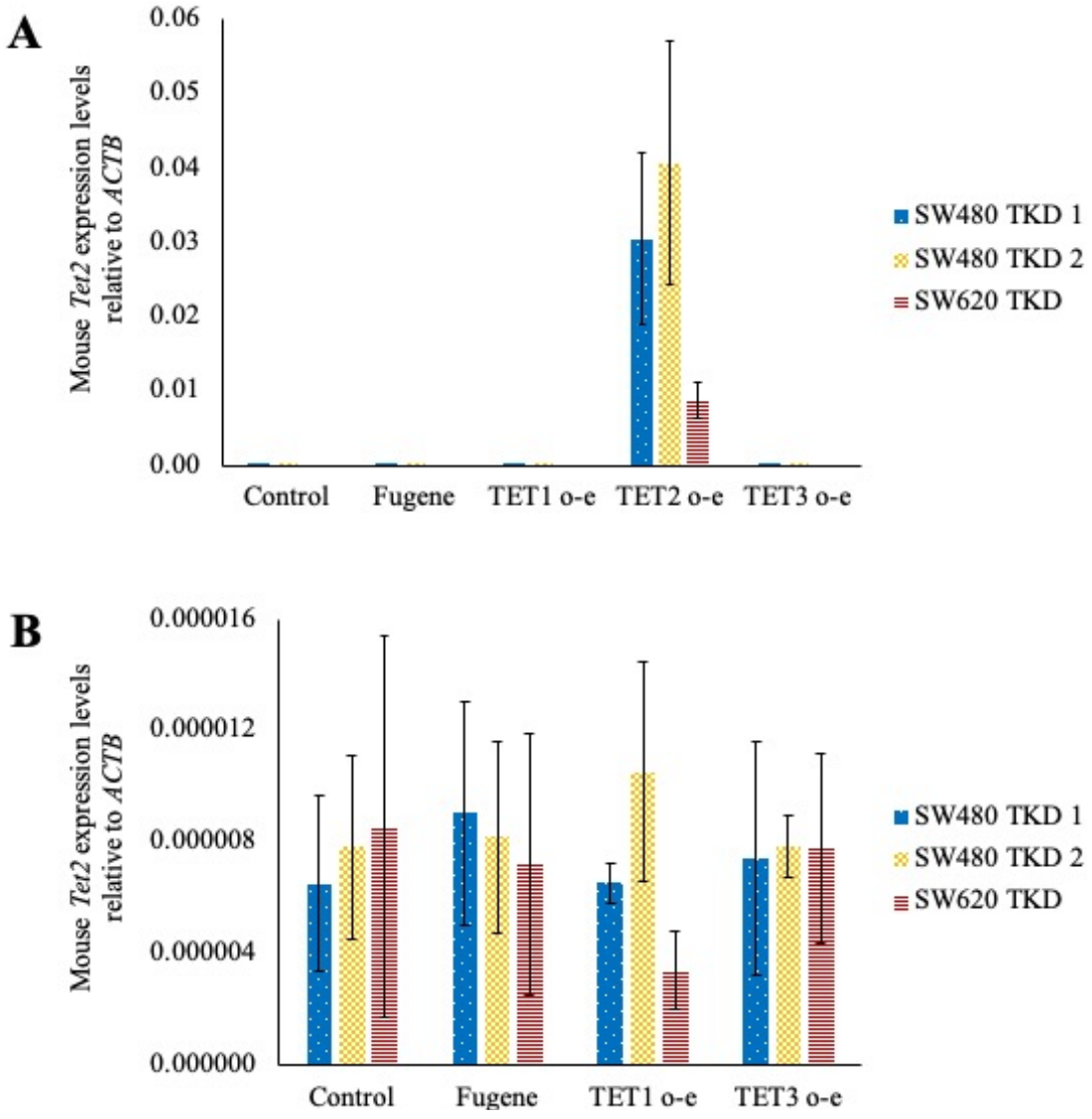
Supplementary Figure 8 The *TET* knockdown SW480 cell lines show a greater reduction in TET activity that the SW620 cell lines compared to the wild type counterparts

The results of the TET activity assay in wild type (WT), single (SKD), double (DKD) and triple (TKD) *TET* knockdown (A) SW480 and (B) SW620 cell lines. The error bars represent the standard deviation of two independent replicates. Only one WT sample for each cell line was included in the results due to an abnormally low reading of the second sample for both cell lines.

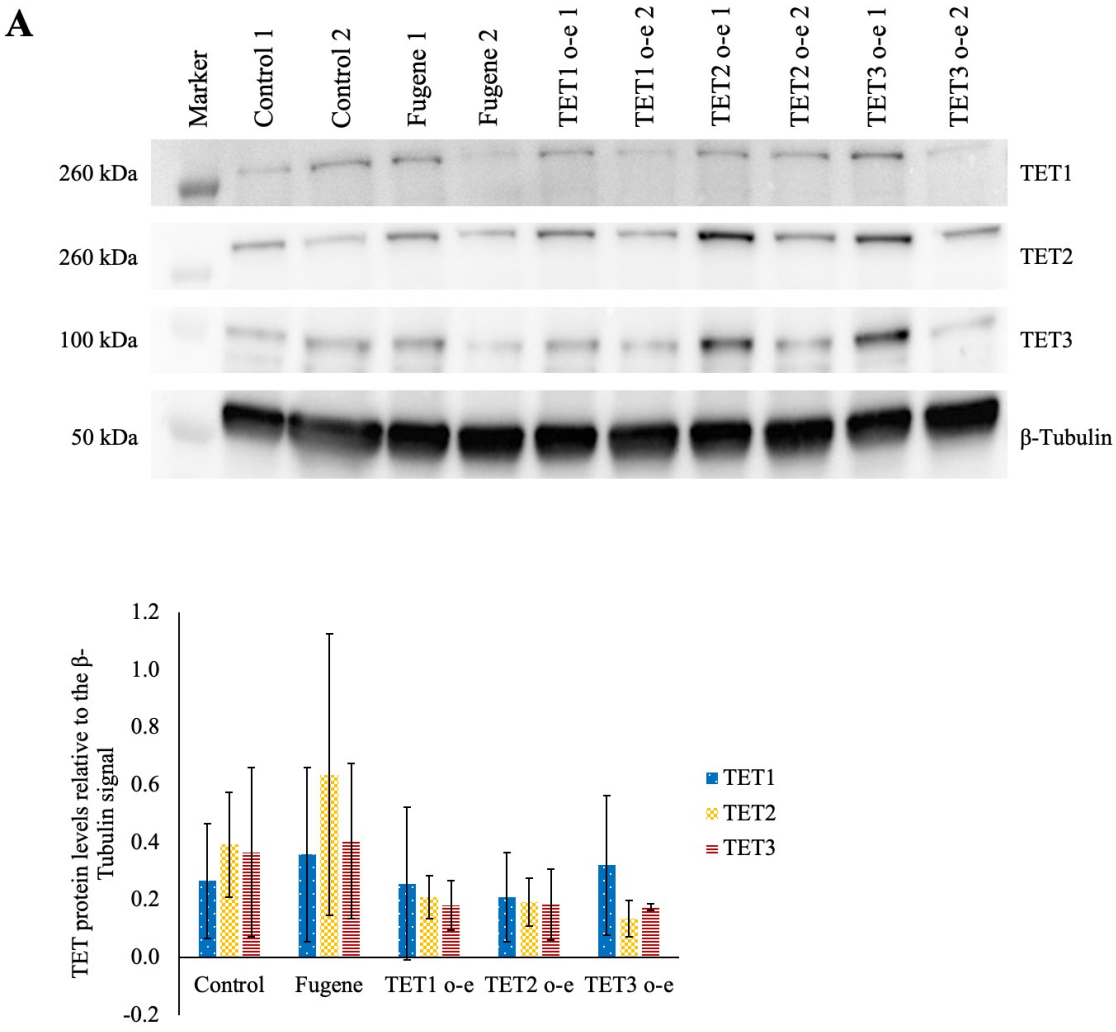


Supplementary Figure 9 The triple *TET* knockdown SW620 cell line has elevated 5mC levels compared to the wild type counterpart

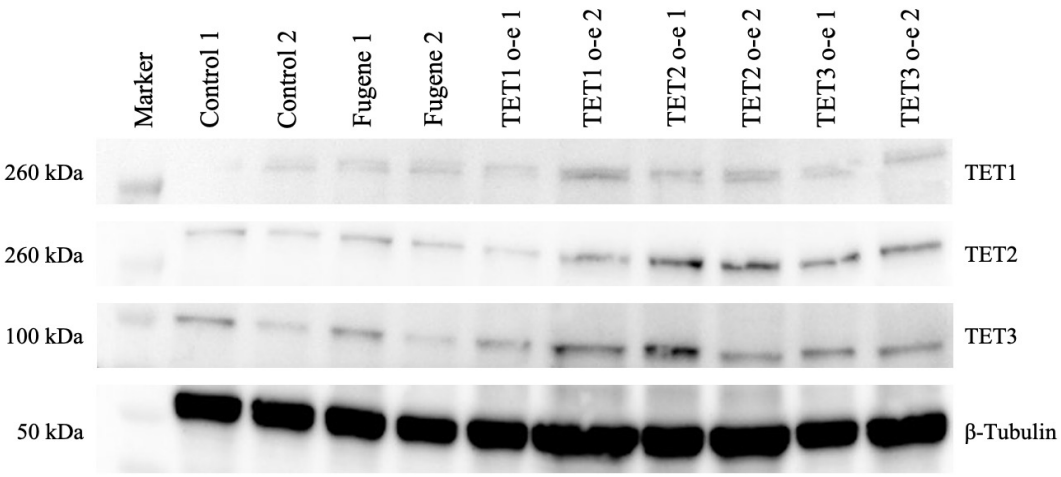
The mass spectrometry results of global 5mC levels quantification in wild type (WT) and triple (TKD) *TET* knockdown (A) SW480 and (B) SW620 cell lines. The signal obtained for 5mC is displayed as a percentage of signal obtained for all cytosines. The error bars represent the standard deviation of (A) four and (B) three independent replicates. The significance was established using (A) a one-way ANOVA followed by the Tukey's single-step multiple comparison; (B) an unpaired t-test (*** $p<0.001$).

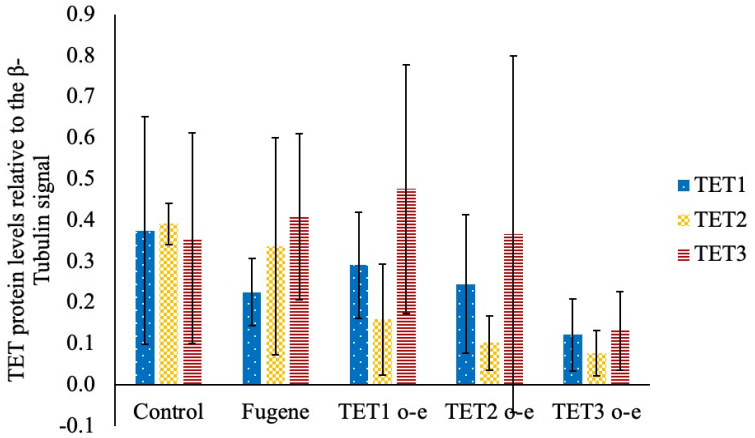


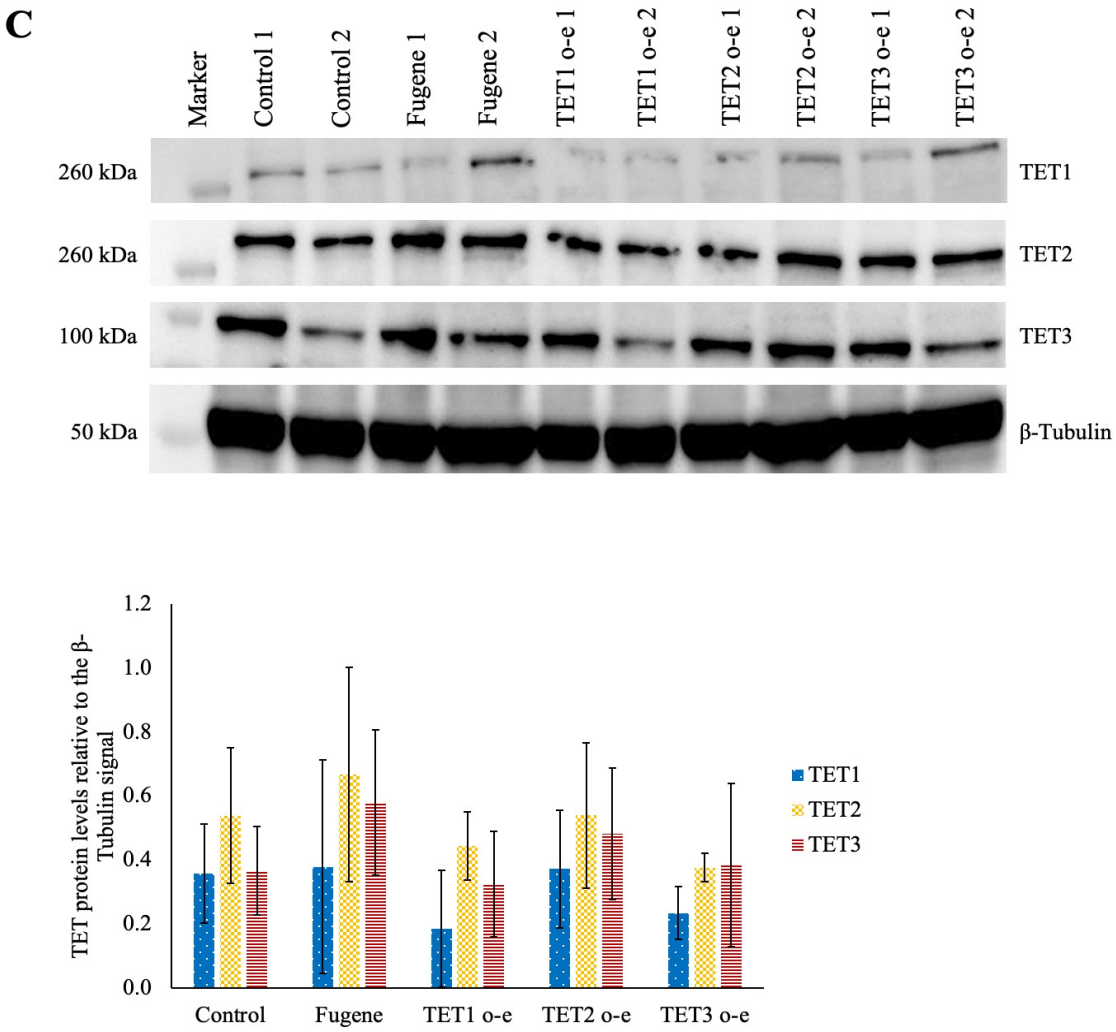
Supplementary Figure 10 The confirmation of mouse *Tet2* overexpression relative to *ACTB* reference gene levels in the three triple *TET* knockdown cell lines
(A) Mouse *Tet2* expression relative to ACTB levels in untransfected control, transfection reagent-only control (Fugene) and TET1/2/3 overexpressing (TET1/2/3 o-e) TKD 1 SW480, TKD 2 SW480 and TKD SW620 cell lines. The results were generated using the ΔCt instead of ΔΔCt values (Livak and Schmittgen, 2001).
(B) The same results as above, but with the exclusion of the TET2 o-e samples.


A
Marker
Control 1
Control 2
Fugene 1
Fugene 2
TET1 o-e 1
TET1 o-e 2
TET2 o-e 1
TET2 o-e 2
TET3 o-e 1
TET3 o-e 2
260 kDa
TET1
260 kDa
TET2
100 kDa
TET3
50 kDa
β-Tubulin
TET protein levels relative to the β-Tubulin signal
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
Control
Fugene
TET1 o-e
TET2 o-e
TET3 o-e
TET1
TET2
TET3

B


Marker
Control 1
Control 2
Fugene 1
Fugene 2
TET1 o-e 1
TET1 o-e 2
TET2 o-e 1
TET2 o-e 2
TET3 o-e 1
TET3 o-e 2
260 kDa
TET1
260 kDa
TET2
100 kDa
TET3
50 kDa
β-Tubulin

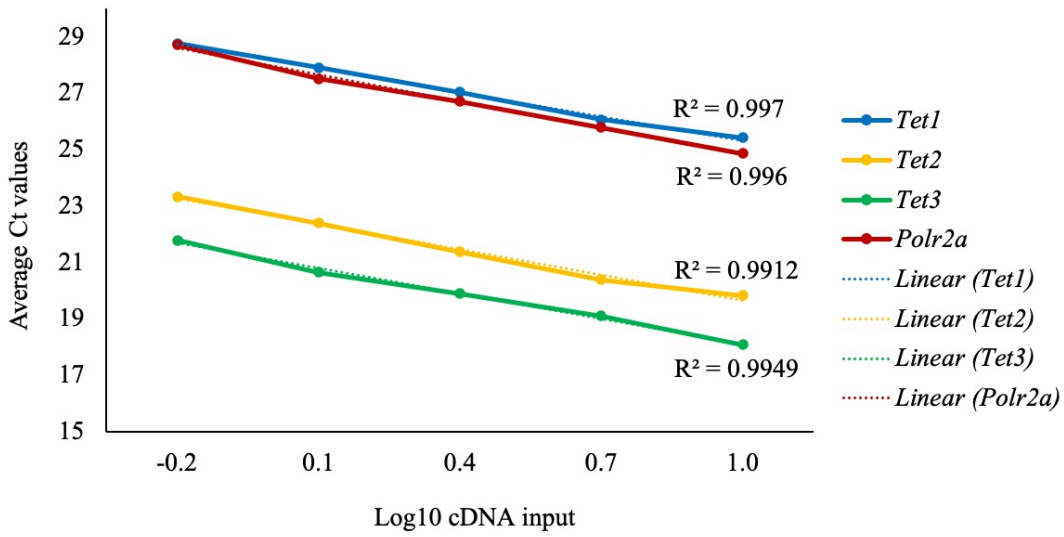

TET protein levels relative to the β-Tubulin signal
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
Control
Fugene
TET1 o-e
TET2 o-e
TET3 o-e
TET1
TET2
TET3



Supplementary Figure 11 The TET protein levels are unchanged by the acute overexpression of the individual TET enzymes

The representative images of Western blots and the graphical representation of the results of the assessment of TET1, TET2, TET3 and ß-Tubulin protein levels in untransfected control, transfection reagent-only control (Fugene) and TET1/2/3 overexpressing (A) TKD 1 SW480, (B) TKD 2 SW480 and (C) TKD SW620 cells. In each blot, two independent replicates of samples are shown, denoted as "1" and "2". The staining results were normalised to the ß-Tubulin content in each cell line. The error bars represent the standard deviation of three independent replicates.

A



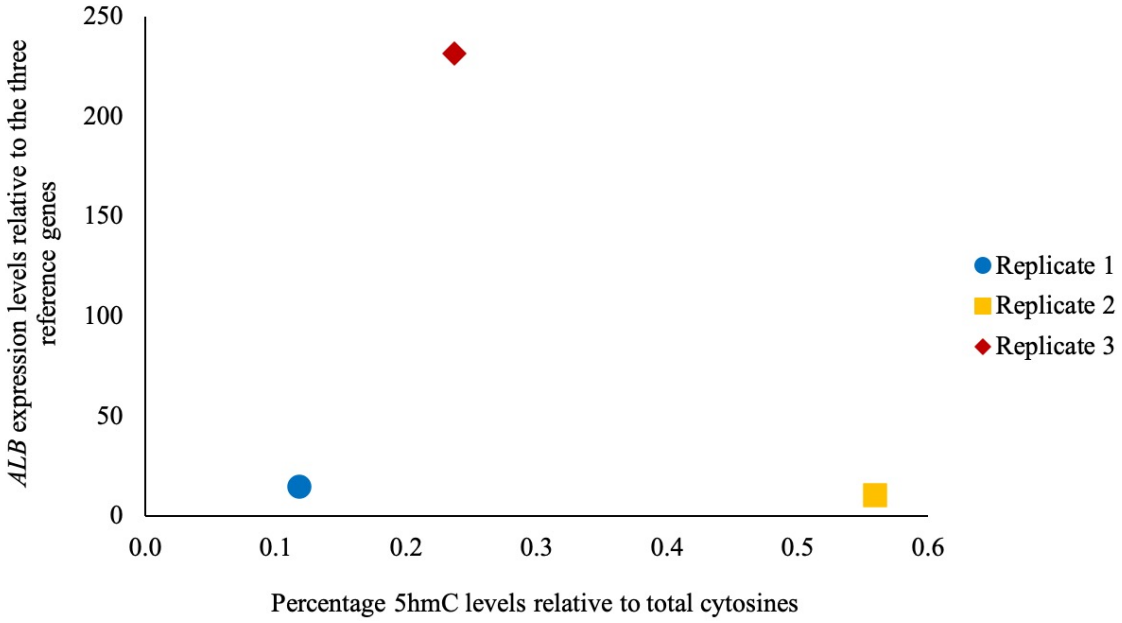
B

Primer pair	Log10 graph slope	Primer efficiency based on the log10 slope (%)	E (for Pfaffl method)
Tet1	-2.84	125	2.25
Tet2	-2.99	116	2.16
Tet3	-2.97	117	2.17
Polr2a	-3.13	109	2.09

Supplementary Figure 12 The experimentally determined primer efficiencies of the *Tet1/2/3* and *Polr2a* primers

(A) The standard curves for the *Tet1/2/3* and *Polr2a* primer pairs depicting the linearity (R^2 values) of amplification efficiency in a 10-fold cDNA dilution series. The dotted lines represent the lines of the best fit.

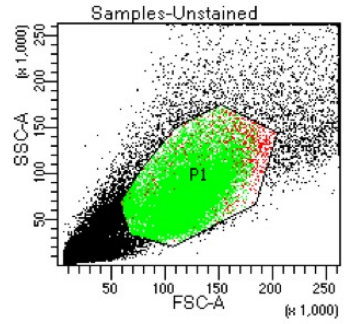
(B) The calculated primer efficiency values for the Pfaffl method (Pfaffl, 2001).

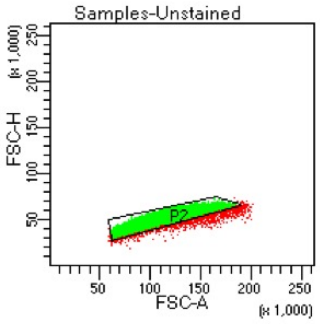


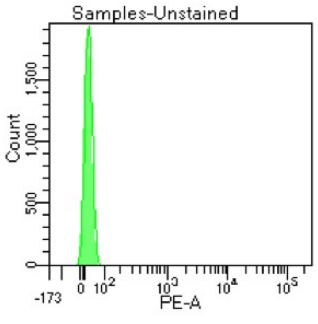
Supplementary Figure 13 The *ALB* expression in hepatocyte-like cells does not show a correlation with their global 5hmC levels

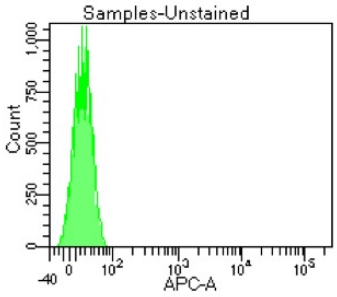
The *ALB* expression relative to the three reference genes used was plotted against the global 5hmC levels for each of the three independent replicates. For each of the three replicates the *ALB* expression and 5hmC levels were analysed in the same sample.

A



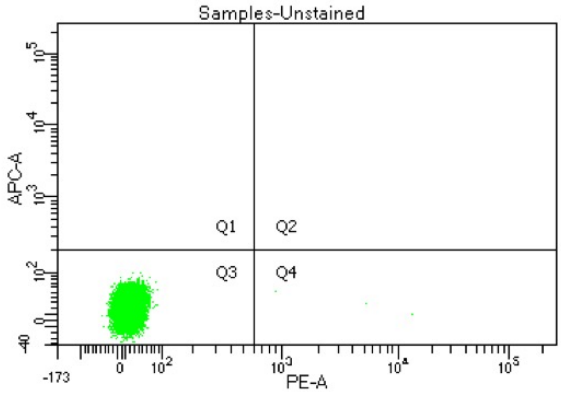


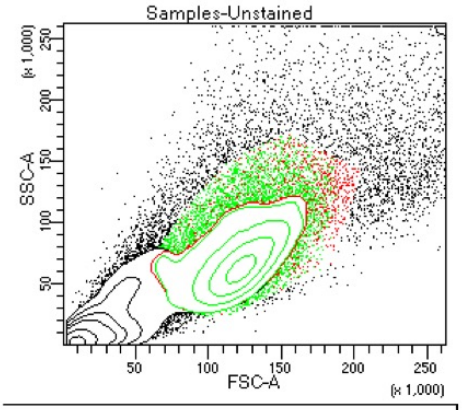




Tube: Unstained

Population	#Events	%Parent	%Total
All Events	50,000	####	100.0
P1	19,478	39.0	39.0
P2	18,043	92.6	36.1
Q1	0	0.0	0.0
Q2	0	0.0	0.0
Q3	18,040	100.0	36.1
Q4	3	0.0	0.0

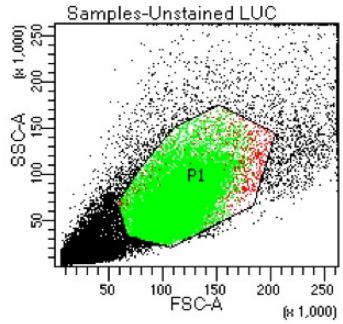


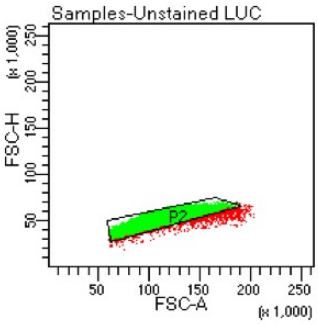


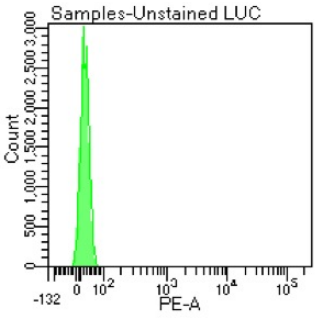
Experiment Name:	20190605
Specimen Name:	Samples
Tube Name:	Unstained
Record Date:	Jun 5, 2019 2:33:28 PM
$OP:	Murrell
GUID:	3c740a60-5d55-405a-a2f9-101f80945111

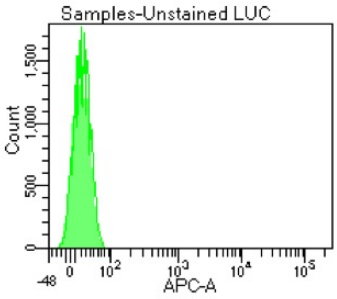
Population	Parent Name	#Events	%Parent	%Grand Par...	%Total	PE-A Mean	APC-A Mean
P2	P1	18,043	92.6	36.1	36.1	14	19
Q1	P2	0	0.0	0.0	0.0	####	####
Q2	P2	0	0.0	0.0	0.0	####	####
Q3	P2	18,040	100.0	92.6	36.1	13	19
Q4	P2	3	0.0	0.0	0.0	6,564	29

B



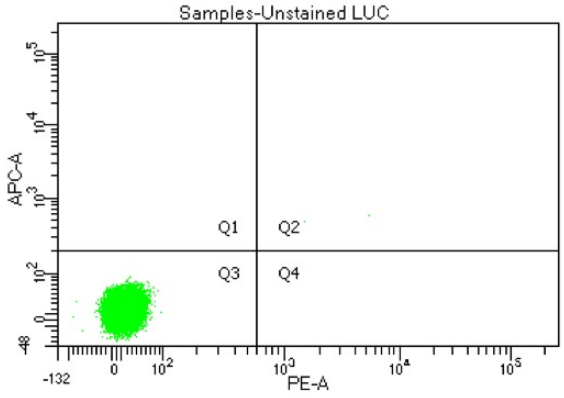


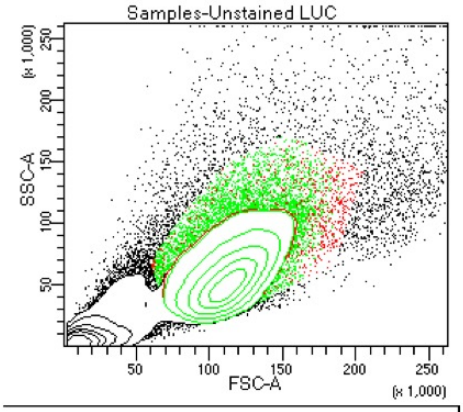




Tube: Unstained LUC

Population	#Events	%Parent	%Total
All Events	50,000	####	100.0
P1	30,152	60.3	60.3
P2	29,031	96.3	58.1
Q1	0	0.0	0.0
Q2	2	0.0	0.0
Q3	29,029	100.0	58.1
Q4	0	0.0	0.0

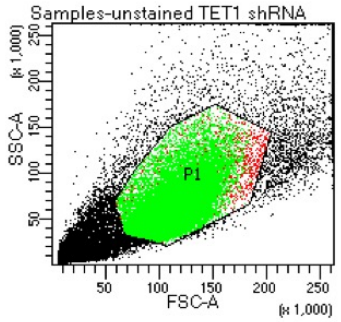


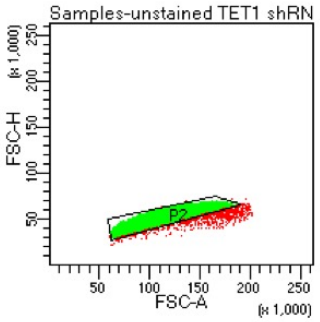


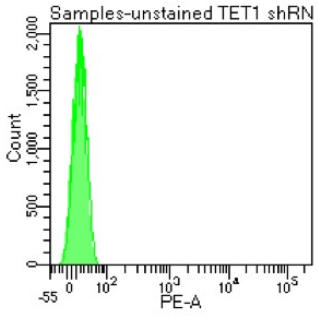
Experiment Name:	20190605
Specimen Name:	Samples
Tube Name:	Unstained LUC
Record Date:	Jun 5, 2019 2:41:29 PM
$OP:	Murrell
GUID:	02087da3-84b4-4c90-9757-b845e4cec4b5

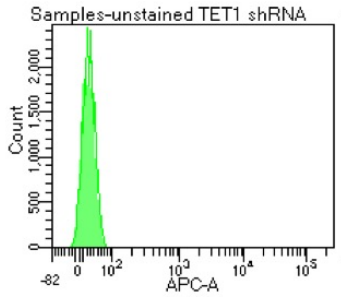
Population	Parent Name	#Events	%Parent	%Grand Par...	%Total	PE-A Mean	APC-A Mean
P2	P1	29,031	96.3	58.1	58.1	12	17
Q1	P2	0	0.0	0.0	0.0	####	####
Q2	P2	2	0.0	0.0	0.0	3,383	523
Q3	P2	29,029	100.0	96.3	58.1	12	17
Q4	P2	0	0.0	0.0	0.0	####	####

C



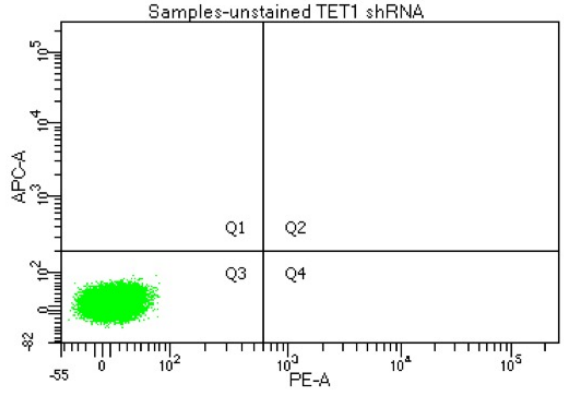


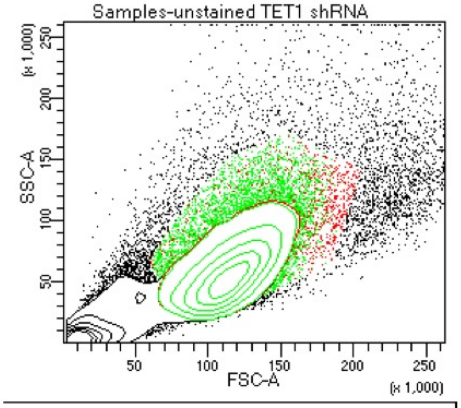




Tube: unstained TET1 shRNA

Population	#Events	%Parent	%Total
All Events	50,000	####	100.0
P1	32,050	64.1	64.1
P2	30,903	96.4	61.8
Q1	0	0.0	0.0
Q2	0	0.0	0.0
Q3	30,903	100.0	61.8
Q4	0	0.0	0.0

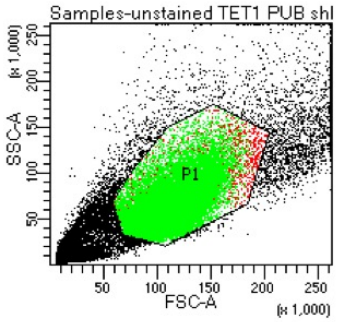


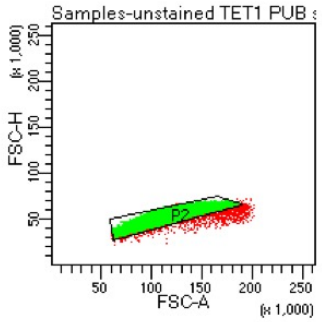


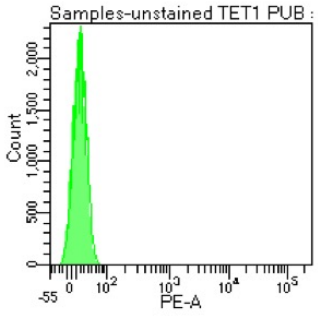
Experiment Name:	20190605
Specimen Name:	Samples
Tube Name:	unstained TET1 shRNA
Record Date:	Jun 5, 2019 2:49:40 PM
$OP:	Murrell
GUID:	3af504e6-451d-4d8f-b265-07c48853f6c9

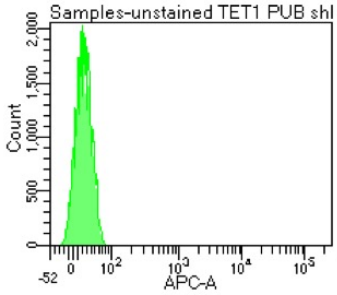
Population	Parent Name	#Events	%Parent	%Grand Par...	%Total	PE-A Mean	APC-A Mean
P2	P1	30,903	96.4	61.8	61.8	12	17
Q1	P2	0	0.0	0.0	0.0	####	####
Q2	P2	0	0.0	0.0	0.0	####	####
Q3	P2	30,903	100.0	96.4	61.8	12	17
Q4	P2	0	0.0	0.0	0.0	####	####

D



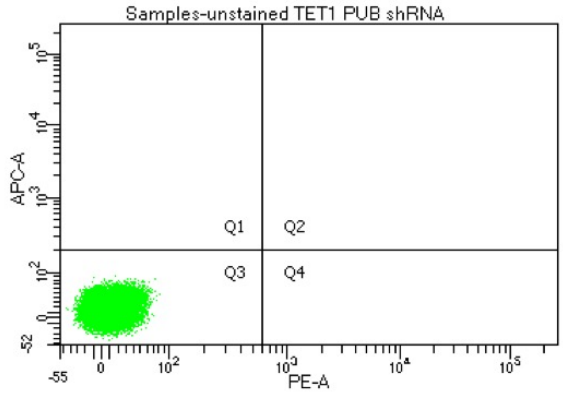


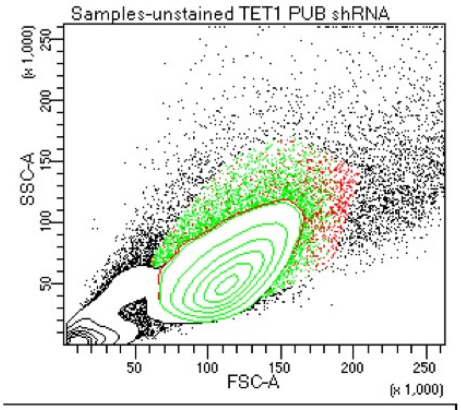




Tube: unstained TET1 PUB shRNA

Population	#Events	%Parent	%Total
All Events	50,000	####	100.0
P1	35,213	70.4	70.4
P2	33,871	96.2	67.7
Q1	0	0.0	0.0
Q2	0	0.0	0.0
Q3	33,871	100.0	67.7
Q4	0	0.0	0.0

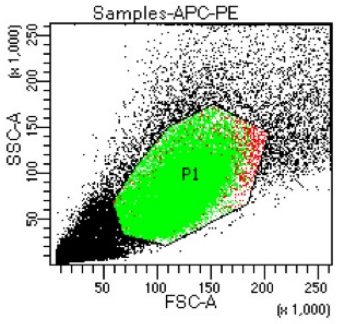


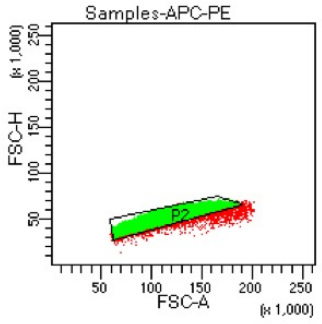


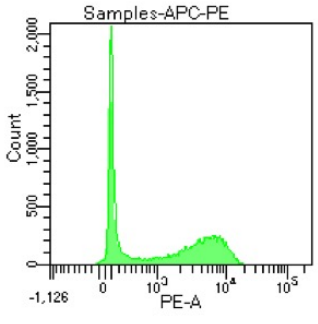
Experiment Name:	20190605
Specimen Name:	Samples
Tube Name:	unstained TET1 PUB shRNA
Record Date:	Jun 5, 2019 2:55:08 PM
$OP:	Murrell
GUID:	e886461a-2f52-42ac-9a58-2dd8ab52f09c

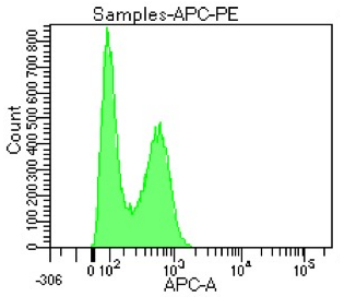
Population	Parent Name	#Events	%Parent	%Grand Par...	%Total	PE-A Mean	APC-A Mean
P2	P1	33,871	96.2	67.7	67.7	12	17
Q1	P2	0	0.0	0.0	0.0	####	####
Q2	P2	0	0.0	0.0	0.0	####	####
Q3	P2	33,871	100.0	96.2	67.7	12	17
Q4	P2	0	0.0	0.0	0.0	####	####

E



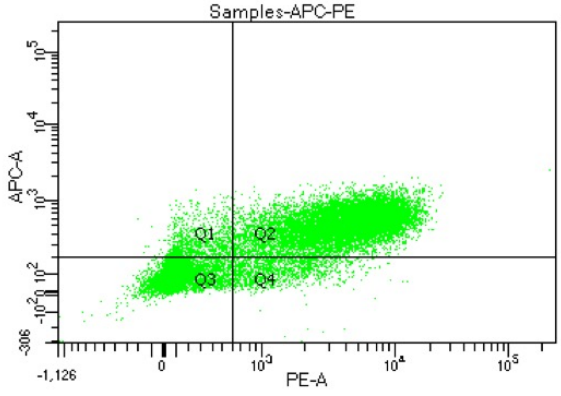


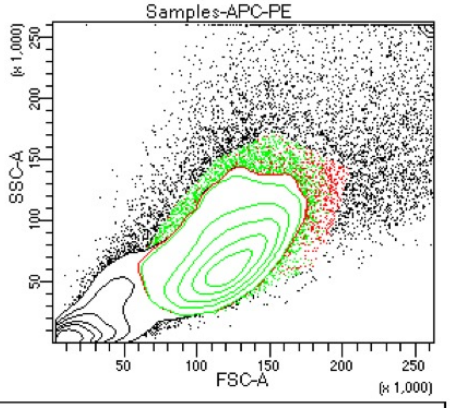




Tube: APC-PE

Population	#Events	%Parent	%Total
All Events	50,000	####	100.0
P1	28,600	57.2	57.2
P2	26,911	94.1	53.8
Q1	993	3.7	2.0
Q2	12,555	46.7	25.1
Q3	12,054	44.8	24.1
Q4	1,309	4.9	2.6

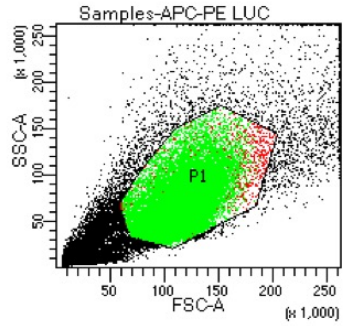


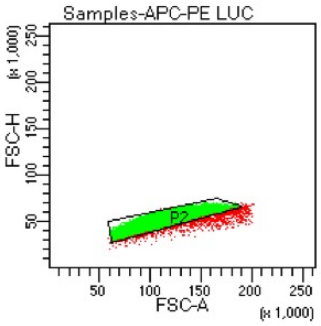


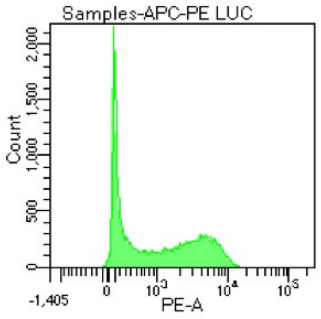
Experiment Name:	20190605
Specimen Name:	Samples
Tube Name:	APC-PE
Record Date:	Jun 5, 2019 2:39:28 PM
$OP:	Murrell
GUID:	bc1501ba-d5e5-433b-b97c-434a95fca151

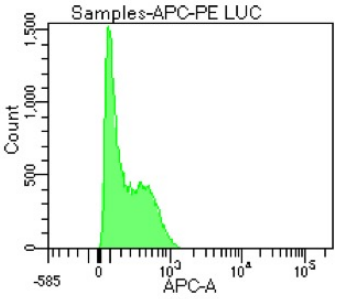
Population	Parent Name	#Events	%Parent	%Grand Par...	%Total	PE-A Mean	APC-A Mean
P2	P1	26,911	94.1	53.8	53.8	2,473	319
Q1	P2	993	3.7	3.5	2.0	269	352
Q2	P2	12,555	46.7	43.9	25.1	5,040	566
Q3	P2	12,054	44.8	42.1	24.1	73	81
Q4	P2	1,309	4.9	4.6	2.6	1,637	119

F



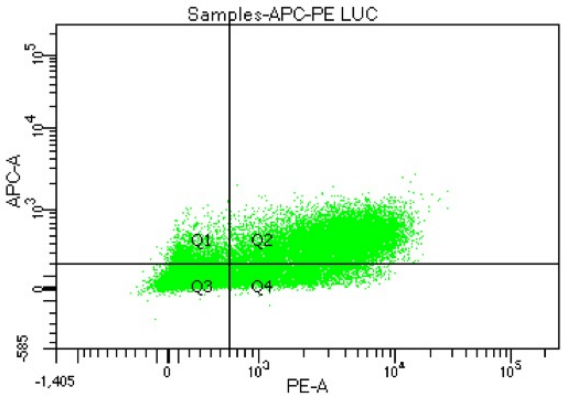


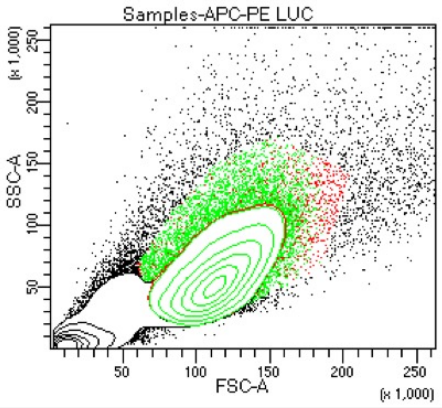




Tube: APC-PE LUC

Population	#Events	%Parent	%Total
All Events	50,000	####	100.0
P1	34,420	68.8	68.8
P2	33,264	96.6	66.5
Q1	1,565	4.7	3.1
Q2	12,751	38.3	25.5
Q3	14,109	42.4	28.2
Q4	4,839	14.5	9.7

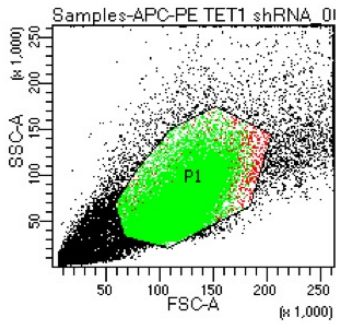


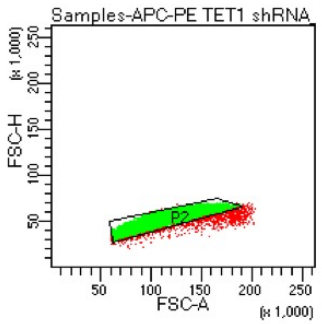


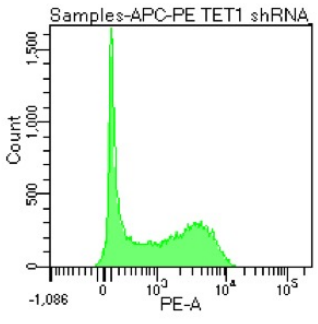
Experiment Name:	20190605
Specimen Name:	Samples
Tube Name:	APC-PE LUC
Record Date:	Jun 5, 2019 2:47:37 PM
$OP:	Murrell
GUID:	83a21037-982c-4bcf-b146-40e5e3b253e8

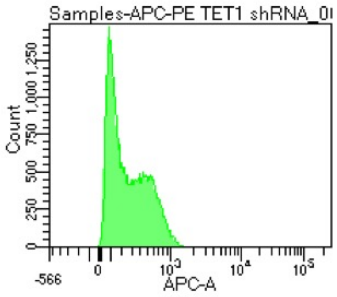
Population	Parent Name	#Events	%Parent	%Grand Par...	%Total	PE-A Mean	APC-A Mean
P2	P1	33,264	96.6	66.5	66.5	1,870	244
Q1	P2	1,565	4.7	4.5	3.1	255	322
Q2	P2	12,751	38.3	37.0	25.5	4,021	472
Q3	P2	14,109	42.4	41.0	28.2	126	76
Q4	P2	4,839	14.5	14.1	9.7	1,811	112

G



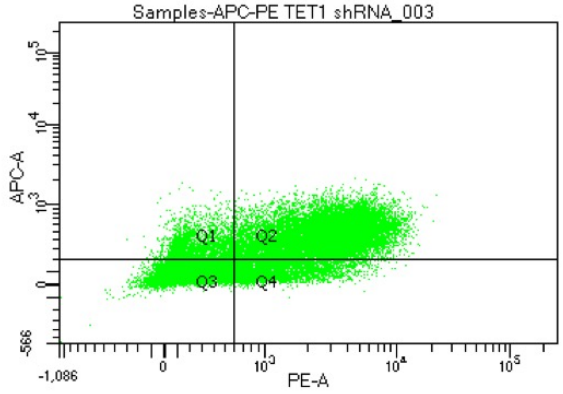


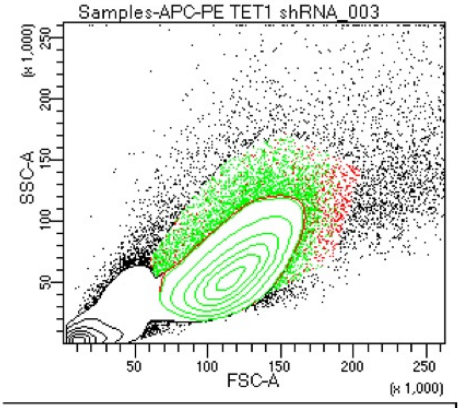




Tube: APC-PE TET1 shRNA_003

Population	#Events	%Parent	%Total
All Events	50,000	####	100.0
P1	36,205	72.4	72.4
P2	35,002	96.7	70.0
Q1	2,554	7.3	5.1
Q2	14,164	40.5	28.3
Q3	14,505	41.4	29.0
Q4	3,779	10.8	7.6

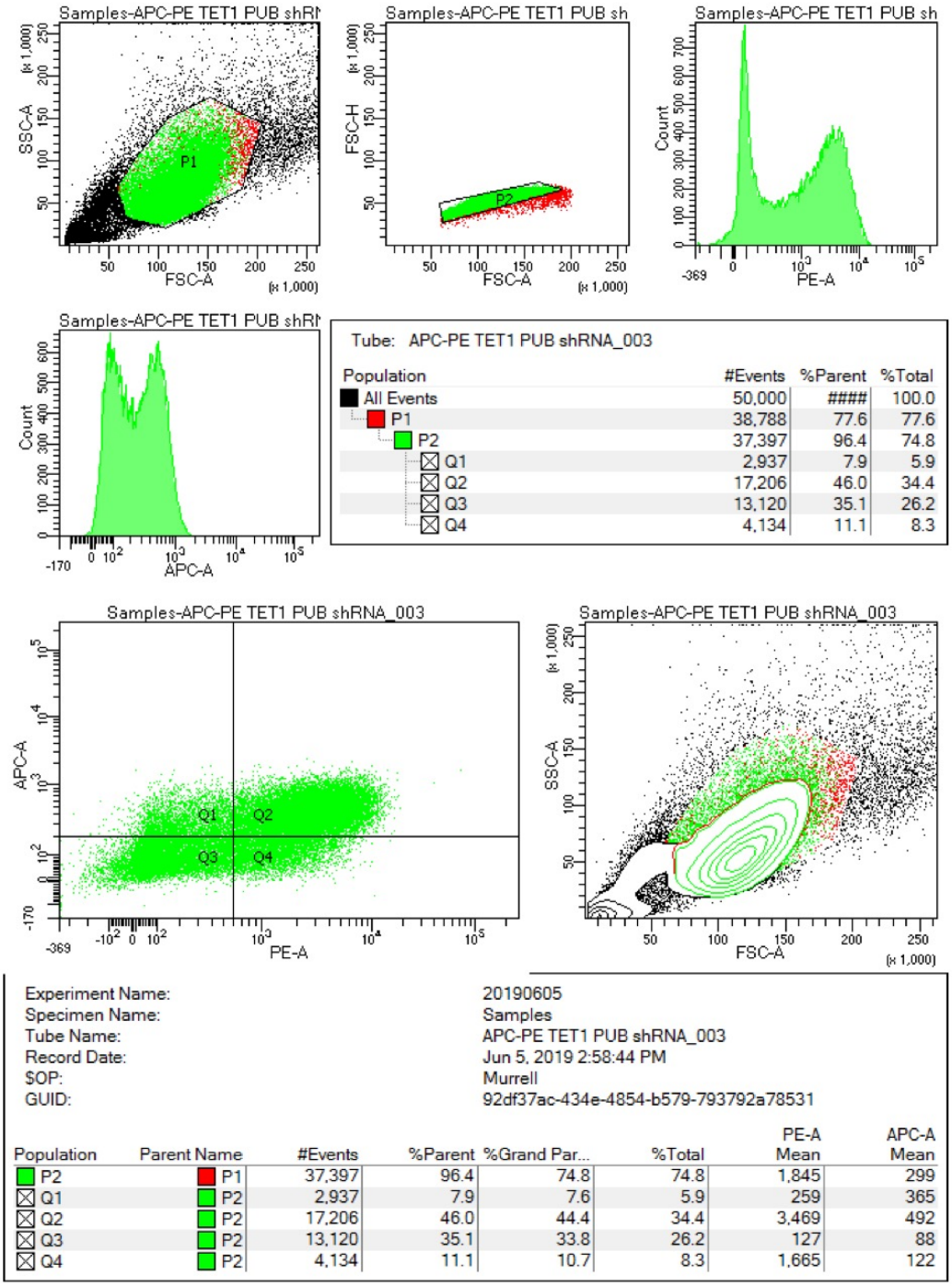




Experiment Name:	20190605
Specimen Name:	Samples
Tube Name:	APC-PE TET1 shRNA_003
Record Date:	Jun 5, 2019 2:53:06 PM
$OP:	Murrell
GUID:	3f4867d0-13e0-40f9-807c-dad1b23e9ca8

Population	Parent Name	#Events	%Parent	%Grand Par...	%Total	PE-A Mean	APC-A Mean
P2	P1	35,002	96.7	70.0	70.0	1,603	269
Q1	P2	2,554	7.3	7.1	5.1	253	344
Q2	P2	14,164	40.5	39.1	28.3	3,365	491
Q3	P2	14,505	41.4	40.1	29.0	123	79
Q4	P2	3,779	10.8	10.4	7.6	1,586	114

H



Experiment Name: 20190605
Specimen Name: Samples
Tube Name: APC-PE TET1 PUB shRNA_003
Record Date: Jun 5, 2019 2:58:44 PM
$OP: Murrell
GUID: 92df37ac-434e-4854-b579-793792a78531

Population	Parent Name	#Events	%Parent	%Grand Par...	%Total	PE-A Mean	APC-A Mean
P2	P1	37,397	96.4	74.8	74.8	1,845	299
Q1	P2	2,937	7.9	7.6	5.9	259	365
Q2	P2	17,206	46.0	44.4	34.4	3,469	492
Q3	P2	13,120	35.1	33.8	26.2	127	88
Q4	P2	4,134	11.1	10.7	8.3	1,665	122

Supplementary Figure 14 The representative examples of FACS analysis results of staining for CD117 and CD184 definitive endoderm markers the in the *TET1* knockdown human embryonic stem cell samples

The details of the FACS sorting for APC (CD117) and PE (CD184) staining in (A-D) unstained and (E-H) stained with both antibodies (A and E) untransfected control, (B and F) luciferase shRNA control, (C and G) TET1 shRNA 1 and (D and H) TET1 shRNA 2 samples. The unstained samples were used to set the background signal for both APC and PE stains.

Reference list

Aapola, U., Shibuya, K., Scott, H.S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Kawasaki, K., Minoshima, S., Krohn, K., Antonarakis, S.E., Shimizu, N., Kudoh, J. and Peterson, P., 2000. Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics*, 65(3), pp.293–298.

Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J.J., Kilpivaara, O., Bhat, R., Huberman, K., Thomas, S., Dolgalev, I., Heguy, A., Paietta, E., Le Beau, M.M., Beran, M., Tallman, M.S., Ebert, B.L., Kantarjian, H.M., Stone, R.M., Gilliland, D.G., Crispino, J.D. and Levine, R.L., 2009. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood*, 114(1), pp.144–147.

Abdouni, H., King, J.J., Suliman, M., Quinlan, M., Fifield, H. and Larijani, M., 2013. Zebrafish AID is capable of deaminating methylated deoxycytidines. *Nucleic Acids Research*, 41(10), pp.5457–5468.

Abedini, P., Fattahi, A., Agah, S., Talebi, A., Beygi, A.H., Amini, S.M., Mirzaei, A. and Akbari, A., 2019. Expression analysis of circulating plasma long noncoding RNAs in colorectal cancer: The relevance of lncRNAs ATB and CCAT1 as potential clinical hallmarks. *Journal of Cellular Physiology*, 234(12), pp.22028–22033.

Advani, S. and Kopetz, S., 2019. Ongoing and future directions in the management of metastatic colorectal cancer: Update on clinical trials. *Journal of Surgical Oncology*, 119(5), pp.642–652.

Advani, S.M., Advani, P., DeSantis, S.M., Brown, D., VonVille, H.M., Lam, M., Loree, J.M., Sarshekeh, A.M., Bressler, J., Lopez, D.S., Daniel, C.R., Swartz, M.D. and Kopetz, S., 2018. Clinical, Pathological, and Molecular Characteristics of CpG Island Methylator Phenotype in Colorectal Cancer: A Systematic Review and Meta-analysis. *Translational Oncology*, 11(5), pp.1188–1201.

Advani, S.M., Advani, P.S., Brown, D.W., Desantis, S.M., Korphaisarn, K., Vonville, H.M., Bressler, J., Lopez, D.S., Davis, J.S., Daniel, C.R., Sarshekeh, A.M., Braithwaite, D., Swartz, M.D. and Kopetz, S., 2019. Global differences in the prevalence of the CpG island methylator phenotype of colorectal cancer. *BMC Cancer*, 19(1), pp.1–13.

Agilent, n.d. *Development of the World's Most Sensitive Triple Quad MS* [Online]. Available from: https://www.agilent.com/cs/library/slidepresentation/Public/6490LCMS_QQQ_presentation.pdf [Accessed 1 May 2020].

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J. and Clarke, M.F., 2003. Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), pp.3983–3988.

Alayed, K., Patel, K.P., Konoplev, S., Singh, R.R., Routbort, M.J., Reddy, N., Pemmaraju, N., Zhang, L., Shaikh, A. Al, Aladily, T.N., Jain, N., Luthra, R., Jeffrey Medeiros, L. and Khoury, J.D., 2013. *TET2* mutations, myelodysplastic features, and a distinct immunoprofile characterize blastic plasmacytoid dendritic cell neoplasm in the bone marrow. *American Journal of Hematology*, 88(12), pp.1055–1061.

Alberge, J.-B., Magrangeas, F., Wagner, M., Denié, S., Guérin-Charbonnel, C., Campion, L., Attal, M., Avet-Loiseau, H., Carell, T., Moreau, P., Minvielle, S. and Sérandour, A., 2019. DNA hydroxymethylation reveals transcription regulation networks and prognostic signatures in multiple myeloma. *bioRxiv*, p.806133.

Alexander, K.A., Wang, X., Shibata, M., Clark, A.G. and García-García, M.J., 2015. TRIM28 Controls Genomic Imprinting through Distinct Mechanisms during and after Early Genome-wide Reprogramming. *Cell Reports*, 13(6), pp.1194–1205.

Amary, M.F., Bacsi, K., Maggiani, F., Damato, S., Halai, D., Berisha, F., Pollock, R., O'Donnell, P., Grigoriadis, A., Diss, T., Eskandarpour, M., Presneau, N., Hogendoorn, P.C.W., Futreal, A., Tirabosco, R. and Flanagan, A.M., 2011. IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours. *Journal of Pathology*, 224(3), pp.334–343.

Ambati, C.R., Vantaku, V., Donepudi, S.R., Amara, C.S., Ravi, S.S., Mandalapu, A., Perla, M., Putluri, V., Sreekumar, A. and Putluri, N., 2019. Measurement of methylated metabolites using liquid chromatography-mass spectrometry and its biological application. *Analytical Methods*, 11(1), pp.49–57.

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W.S., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., Encheva, V., Kudo, N.R., Koseki, H., Sasaki, H. and Hajkova, P., 2016. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nature Cell Biology*, 18(2), pp.225–233.

An, J., González-Avalos, E., Chawla, A., Jeong, M., López-Moyado, I.F., Li, W., Goodell, M.A., Chavez, L., Ko, M. and Rao, A., 2015. Acute loss of TET function results in aggressive myeloid cancer in mice. *Nature Communications*, 6.

Annesley, T.M., 2003. Ion Suppression in Mass Spectrometry. *Clinical Chemistry*, 49(7), pp.1041–1044.

Araghi, M., Soerjomataram, I., Bardot, A., Ferlay, J., Cabasag, C.J., Morrison, D.S., De, P., Tervonen, H., Walsh, P.M., Bucher, O., Engholm, G., Jackson, C., McClure, C., Woods, R.R., Saint-Jacques, N., Morgan, E., Ransom, D., Thursfield, V., Møller, B., Leonfellner, S., Guren, M.G., Bray, F. and Arnold, M., 2019. Changes in colorectal cancer incidence in seven high-income countries: a population-based study. *The Lancet Gastroenterology and Hepatology*, 4(7), pp.511–518.

Arai, M., Nobusawa, S., Ikota, H., Takemura, S. and Nakazato, Y., 2012. Frequent IDH1/2 mutations in intracranial chondrosarcoma: A possible diagnostic clue for its differentiation from chordoma. *Brain Tumor Pathology*, 29(4), pp.201–206.

Arnold, M., Sierra, M.S., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F., 2017. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*, 66(4), pp.683–691.

Ashapkin, V. V., Kutueva, L.I. and Vanyushin, B.F., 2016. Dnmt2 is the most evolutionary conserved and enigmatic cytosine DNA methyltransferase in eukaryotes. *Russian Journal of Genetics*, 52(3), pp.237–248. Maik Nauka Publishing / Springer SBM.

Aubrey, B.J., Kelly, G.L., Janic, A., Herold, M.J. and Strasser, A., 2018. How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? *Cell Death and Differentiation*, 25(1), pp.104–113. Nature Publishing Group.

Avery, O.T., Macleod, C.M. and McCarty, M., 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of

transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. *Journal of Experimental Medicine*, 79(2), pp.137–158.

Azad, N.S., Shirai, K., McRee, A.J., Opyrchal, M., Johnson, D.B., Ordentlich, P., Brouwer, S., Sankoh, S., Schmidt, E. V., Meyers, M.L. and Johnson, M.L., 2018. ENCORE 601: A phase 2 study of entinostat in combination with pembrolizumab in patients with microsatellite stable metastatic colorectal cancer. *Journal of Clinical Oncology*, 36(15), pp.3557–3557.

Ba, X. and Boldogh, lstvan, 2018. 8-Oxoguanine DNA glycosylase 1: Beyond repair of the oxidatively modified base lesions. *Redox Biology*, 14, pp.669–678. Elsevier B.V.

Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H.E., Iurlaro, M., Reik, W., Murrell, A. and Balasubramanian, S., 2015. 5-Formylcytosine can be a stable DNA modification in mammals. *Nature Chemical Biology*, 11(8), pp.555–557.

Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A. and Balasubramanian, S., 2014. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nature Chemistry*, 6(12), pp.1049–1055.

Badal, V., Chuang, L.S.H., Tan, E.H.-H., Badal, S., Villa, L.L., Wheeler, C.M., Li, B.F.L. and Bernard, H.-U., 2003. CpG Methylation of Human Papillomavirus Type 16 DNA in Cervical Cancer Cell Lines and in Clinical Specimens: Genomic Hypomethylation Correlates with Carcinogenic Progression. *Journal of Virology*, 77(11), pp.6227–6234.

Balasubramani, A. and Rao, A., 2013. O-GlcNAcylation and 5-Methylcytosine Oxidation: An Unexpected Association between OGT and TETs. *Molecular Cell*, 49(4), pp.618–619. Cell Press.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., Leproust, E.M., Park, I.H., Xie, B., Daley, G.Q. and Church, G.M., 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature Biotechnology*, 27(4), pp.361–368.

Banerjee, S. and Mazumdar, S., 2012. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry*, 2012, p.40.

Bannister, A.J. and Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell Research*, 21(3), pp.381–395. Nature Publishing Group.

Barazeghi, E., Gill, A.J., Sidhu, S., Norlen, O., Dina, R., Palazzo, F.F., Hellman, P., Stalberg, P. and Westin, G., 2017. A role for TET2 in parathyroid carcinoma. *Endocrine-Related Cancer*, 24(7), pp.329–338.

Barker, N., Van Es, J.H., Kuipers, J., Kujala, P., Van Den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J. and Clevers, H., 2007. Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature*, 449(7165), pp.1003–1007.

Barker, N., Van Oudenaarden, A. and Clevers, H., 2012. Identifying the stem cell of the intestinal crypt: Strategies and pitfalls. *Cell Stem Cell*, 11(4), pp.452–460. Cell Press.

Barker, N., Ridgway, R.A., Van Es, J.H., Van De Wetering, M., Begthel, H., Van Den Born, M., Danenberg, E., Clarke, A.R., Sansom, O.J. and Clevers, H., 2009. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature*, 457(7229), pp.608–611.

Bartlett, D.C., Hodson, J., Bhogal, R.H., Youster, J. and Newsome, P.N., 2014. Combined use of N-acetylcysteine and Liberase improves the viability and metabolic function of human hepatocytes isolated from human liver. *Cytotherapy*, 16(6), pp.800–809.

Bauer, C., Göbel, K., Nagaraj, N., Colantuoni, C., Wang, M., Müller, U., Kremmer, E., Rottach, A. and Leonhardt, H., 2015. Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-Linked N-Acetylglucosamine transferase (OGT). *Journal of Biological Chemistry*, 290(8), pp.4801–4812.

Baxter, M., Withey, S., Harrison, S., Segeritz, C.P., Zhang, F., Atkinson-Dell, R., Rowe, C., Gerrard, D.T., Sison-Young, R., Jenkins, R., Henry, J., Berry, A.A., Mohamet, L., Best, M., Fenwick, S.W., Malik, H., Kitteringham, N.R., Goldring, C.E., Piper Hanley, K., Vallier, L. and Hanley, N.A., 2015. Phenotypic and functional analyses show stem cell-derived hepatocyte-like cells better mimic fetal rather than adult hepatocytes. *Journal of Hepatology*, 62(3), pp.581–589.

Benetatos, L. and Vartholomatos, G., 2015. Imprinted genes in myeloid lineage commitment in normal and malignant hematopoiesis. *Leukemia*, 29(6), pp.1233–1242. Nature Publishing Group.

Berger, Y., Cui, X.-L., West, D., Rivas, M., Hsu, P., Dougherty, U., Deng, Z., Hindi, E., Eng, O., Bissonnette, M., Polite, B.N. and Turaga, K., 2020. 5-hydroxymethylation signatures in plasma circulating cell-free DNA as markers for appendiceal and colorectal peritoneal metastasis. *Journal of Clinical Oncology*, 38(4), pp.195–195.

Bestor, T., Laudano, A., Mattaliano, R. and Ingram, V., 1988. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of Molecular Biology*, 203(4), pp.971–983.

Bestor, T.H. and Ingram, V.M., 1983. Two DNA methyltransferases from murine erythroleukemia cells: Purification, sequence specificity, and mode of interaction with DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 80(181), pp.5559–5563.

Beumer, J. and Clevers, H., 2016. Regulation and plasticity of intestinal stem cells during homeostasis and regeneration. *Development*, 143(20), pp.3639–3649. Company of Biologists Ltd.

Bhat, S., Kabekkodu, S.P., Varghese, V.K., Chakrabarty, S., Mallya, S.P., Rotti, H., Pandey, D., Kushtagi, P. and Satyamoorthy, K., 2017. Aberrant gene-specific DNA methylation signature analysis in cervical cancer. *Tumor Biology*, 39(3), p.101042831769457.

Bhattacharyya, S., Pradhan, K., Campbell, N., Mazdo, J., Vasantkumar, A., Maqbool, S., Bhagat, T.D., Gupta, S., Suzuki, M., Yu, Y., Greally, J.M., Steidl, U., Bradner, J., Dawlaty, M., Godley, L., Maitra, A. and Verma, A., 2017. Altered hydroxymethylation is seen at regulatory regions in pancreatic cancer and regulates oncogenic pathways. *Genome Research*, 27(11), pp.1830–1842.

Bhattacharyya, S., Yu, Y., Suzuki, M., Campbell, N., Mazdo, J., Vasanthakumar, A., Bhagat, T.D., Nischal, S., Christopeit, M., Parekh, S., Steidl, U., Godley, L., Maitra, A., Greally, J.M. and Verma, A., 2013. Genome-wide hydroxymethylation tested using the HELP-GT assay shows redistribution in cancer. *Nucleic Acids*

Research, 41(16), pp.e157–e157.

Bilyard, M.K., Becker, S. and Balasubramanian, S., 2020. Natural, modified DNA bases. *Current Opinion in Chemical Biology*, 57, pp.1–7. Elsevier Ltd.

Bird, A., Taggart, M., Frommer, M., Miller, O.J. and Macleod, D., 1985. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell*, 40(1), pp.91–99.

Bird, A.P., 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Research*, 8(7), pp.1499–1504.

Blanco, E., González-Ramírez, M., Alcaine-Colet, A., Aranda, S. and Di Croce, L., 2020. The Bivalent Genome: Characterization, Structure, and Regulation. *Trends in Genetics*, 36(2), pp.118–131. Elsevier Ltd.

Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martínez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., Lorincz, M.C. and Ramalho-Santos, M., 2013. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature*, 500(7461), pp.222–226.

De Both, N.J., Vermey, M., Dinjens, W.N. and Bosman, F.T., 1999. A comparative evaluation of various invasion assays testing colon carcinoma cell lines. *British Journal of Cancer*, 81(6), pp.934–941.

Bott, A.J., Shen, J., Tonelli, C., Zhan, L., Sivaram, N., Jiang, Y.P., Yu, X., Bhatt, V., Chiles, E., Zhong, H., Maimouni, S., Dai, W., Velasquez, S., Pan, J.A., Muthalagu, N., Morton, J., Anthony, T.G., Feng, H., Lamers, W.H., Murphy, D.J., Guo, J.Y., Jin, J., Crawford, H.C., Zhang, L., White, E., Lin, R.Z., Su, X., Tuveson, D.A. and Zong, W.X., 2019. Glutamine Anabolism Plays a Critical Role in Pancreatic Cancer by Coupling Carbon and Nitrogen Metabolism. *Cell Reports*, 29(5), pp.1287-1298.e6.

Bransteitter, R., Pham, P., Scharfft, M.D. and Goodman, M.F., 2003. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), pp.4102–4107.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A., 2018.

Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), pp.394–424.

Brioude, F., Kalish, J.M., Mussa, A., Foster, A.C., Bliek, J., Ferrero, G.B., Boonen, S.E., Cole, T., Baker, R., Bertoletti, M., Cocchi, G., Coze, C., De Pellegrin, M., Hussain, K., Ibrahim, A., Kilby, M.D., Krajewska-Walasek, M., Kratz, C.P., Ladusans, E.J., Lapunzina, P., Bouc, Y. Le, Maas, S.M., MacDonald, F., Õunap, K., Peruzzi, L., Rossignol, S., Russo, S., Shipster, C., Skórka, A., Tatton-Brown, K., Tenorio, J., Tortora, C., Grønskov, K., Netchine, I., Hennekam, R.C., Prawitt, D., Tümer, Z., Eggermann, T., Mackay, D.J.G., Riccio, A. and Maher, E.R., 2018. Clinical and molecular diagnosis, screening and management of Beckwith-Wiedemann syndrome: An international consensus statement. *Nature Reviews Endocrinology*, 14(4), pp.229–249.

Brody, H., 2015. Colorectal cancer. *Nature*, 521(7551), p.S1. Nature Publishing Group.

Buiting, K., Williams, C. and Horsthemke, B., 2016. Angelman syndrome-insights into a rare neurogenetic disorder. *Nature Reviews Neurology*, 12(10), pp.584–593. Nature Publishing Group.

Bulterijs, S., Hull, R.S., Björk, V.C.E. and Roy, A.G., 2015. It is time to classify biological aging as a disease. *Frontiers in Genetics*, 6(JUN), p.205.

Burr, S., Caldwell, A., Chong, M., Beretta, M., Metcalf, S., Hancock, M., Arno, M., Balu, S., Kropf, V.L., Mistry, R.K., Shah, A.M., Mann, G.E. and Brewer, A.C., 2018. Oxygen gradients can determine epigenetic asymmetry and cellular differentiation via differential regulation of Tet activity in embryonic stem cells. *Nucleic Acids Research*, 46(3), pp.1210–1226.

Byun, H.M., Wong, H.L., Birnstein, E.A., Wolff, E.M., Liang, G. and Yang, A.S., 2007. Examination of IGF2 and H19 loss of imprinting in bladder cancer. *Cancer Research*, 67(22), pp.10753–10758.

Calvanese, V., Fernández, A.F., Urdinguio, R.G., Suárez-Alvarez, B., Mangas, C., Pérez-García, V., Bueno, C., Montes, R., Ramos-Mejía, V., Martínez-Camblor, P., Ferrero, C., Assenov, Y., Bock, C., Menendez, P., Carrera, A.C., Lopez-Larrea, C. and Fraga, M.F., 2012. A promoter DNA demethylation landscape of human

hematopoietic differentiation. *Nucleic Acids Research*, 40(1), pp.116–131.

Cameron, E. and Pauling, L., 1976. Supplemental ascorbate in the supportive treatment of cancer: Prolongation of survival times in terminal human cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 73(10), pp.3685–3689.

Cameron, K., Lucendo-Villarin, B., Szkolnicka, D. and Hay, D.C., 2014. Serum-free directed differentiation of human embryonic stem cells to hepatocytes. *Methods in Molecular Biology*, 1250, pp.105–111.

Cao, T., Pan, W., Sun, X. and Shen, H., 2019. Increased expression of TET3 predicts unfavorable prognosis in patients with ovarian cancer-a bioinformatics integrative analysis. *Journal of Ovarian Research*, 12(1), pp.1–10. BioMed Central Ltd.

Cardoso, M.C. and Leonhardt, H., 1999. DNA methyltransferase is actively retained in the cytoplasm during early development. *Journal of Cell Biology*, 147(1), pp.25–32.

Carella, A., Tejedor, J.R., García, M.G., Urdinguio, R.G., Bayón, G.F., Sierra, M., López, V., García-Toraño, E., Santamarina-Ojeda, P., Pérez, R.F., Bigot, T., Mangas, C., Corte-Torres, M.D., Sáenz-de-Santa-María, I., Mollejo, M., Meléndez, B., Astudillo, A., Chiara, M.D., Fernández, A.F. and Fraga, M.F., 2020. Epigenetic downregulation of TET3 reduces genome-wide 5hmC levels and promotes glioblastoma tumorigenesis. *International Journal of Cancer*, 146(2), pp.373–387.

Carey, B.W., Finley, L.W.S., Cross, J.R., Allis, C.D. and Thompson, C.B., 2015. Intracellular α-ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature*, 518(7539), pp.413–416.

Carr, A.C. and Cook, J., 2018. Intravenous vitamin C for cancer therapy - Identifying the current gaps in our knowledge. *Frontiers in Physiology*, 9(AUG), p.1182. Frontiers Media S.A.

Castro, W., Chelbi, S.T., Niogret, C., Ramon-Barros, C., Welten, S.P.M., Osterheld, K., Wang, H., Rota, G., Morgado, L., Vivier, E., Raeber, M.E., Boyman, O., Delorenzi, M., Barras, D., Ho, P.C., Oxenius, A. and Guarda, G., 2018. The transcription factor Rfx7 limits metabolism of NK cells and promotes their

maintenance and immunity. *Nature Immunology*, 19(8), pp.809–820.

Cekanova, M. and Rathore, K., 2014. Animal models and therapeutic molecular targets of cancer: Utility and limitations. *Drug Design, Development and Therapy*, 8, pp.1911–1922. Dove Medical Press Ltd.

Cervena, K., Siskova, A., Buchler, T., Vodicka, P. and Vymetalkova, V., 2020. Methylation-Based Therapies for Colorectal Cancer. *Cells*, 9(6), p.1540. NLM (Medline).

Céspedes, M.V., Espina, C., García-Cabezas, M.A., Trias, M., Boluda, A., Gómez Del Pulgar, M.T., Sancho, F.J., Nistal, M., Lacal, J.C. and Mangues, R., 2007. Orthotopic microinjection of human colon cancer cells in nude mice induces tumor foci in all clinically relevant metastatic sites. *American Journal of Pathology*, 170(3), pp.1077–1085.

Cha, J., Roomi, M.W., Ivanov, V., Kalinovsky, T., Niedzwiecki, A. and Rath, M., 2013. Ascorbate supplementation inhibits growth and metastasisof B16FO melanoma and 4T1 breast cancer cellsin vitamin C-deficient mice. *International Journal of Oncology*, 42(1), pp.55–64.

Chang, J., Chance, M.R., Nicholas, C., Ahmed, N., Guilmeau, S., Flandez, M., Wang, D., Byun, D.S., Nasser, S., Albanese, J.M., Corner, G.A., Heerdt, B.G., Wilson, A.J., Augenlicht, L.H. and Mariadason, J.M., 2008. Proteomic changes during intestinal cell maturation in vivo. *Journal of Proteomics*, 71(5), pp.530–546.

Chang, Y.S., Huang, H. Da, Yeh, K.T. and Chang, J.G., 2016. Genetic alterations in endometrial cancer by targeted next-generation sequencing. *Experimental and Molecular Pathology*, 100(1), pp.8–12.

Chapel, D.B., Husain, A.N. and Krausz, T., 2019. Immunohistochemical evaluation of nuclear 5-hydroxymethylcytosine (5-hmC) accurately distinguishes malignant pleural mesothelioma from benign mesothelial proliferations. *Modern Pathology*, 32(3), pp.376–386.

Chapman, C.G., Mariani, C.J., Wu, F., Meckel, K., Butun, F., Chuang, A., Madzo, J., Bissonette, M.B., Kwon, J.H. and Godley, L.A., 2015. TET-catalyzed 5-hydroxymethylcytosine regulates gene expression in differentiating colonocytes

and colon cancer. *Scientific Reports*, 5(1), p.17568.

Chédin, F., Lieber, M.R. and Hsieh, C.L., 2002. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26), pp.16916–16921.

Chen, D., Maruschke, M., Hakenberg, O., Zimmermann, W., Stief, C.G. and Buchner, A., 2017. TOP2A, HELLS, ATAD2, and TET3 Are Novel Prognostic Markers in Renal Cell Carcinoma. *Urology*, 102, pp.265.e1-265.e7.

Chen, D., Sun, N., Hou, L., Kim, R., Faith, J., Aslanyan, M., Tao, Y., Zheng, Y., Fu, J., Liu, W., Kellis, M. and Clark, A., 2019. Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors. *Cell Reports*, 29(13), pp.4568-4582.e5.

Chen, H., Yu, D., Fang, R., Rabidou, K., Wu, D., Hu, D., Jia, P., Zhao, Z., Wu, Z., Peng, J., Shi, Y. and Shi, Y.G., 2019. TET2 stabilization by 14-3-3 binding to the phosphorylated Serine 99 is deregulated by mutations in cancer. *Cell Research*, 29(3), pp.248–250. Nature Publishing Group.

Chen, J., Guo, L., Zhang, L., Wu, Haoyu, Yang, J., Liu, H., Wang, X., Hu, X., Gu, T., Zhou, Z., Liu, Jing, Liu, Jiadong, Wu, Hongling, Mao, S.Q., Mo, K., Li, Y., Lai, K., Qi, J., Yao, H., Pan, G., Xu, G.L. and Pei, D., 2013. Vitamin C modulates TET1 function during somatic cell reprogramming. *Nature Genetics*, 45(12), pp.1504–1509.

Chen, J., Wu, L., Xu, H. and Cheng, S., 2019. 5-Aza-CdR regulates RASSF1A by inhibiting DNMT1 to affect colon cancer cell proliferation, migration and apoptosis. *Cancer Management and Research*, 11, pp.9517–9528.

Chen, K., Zhang, J., Guo, Z., Ma, Q., Xu, Zhengzheng, Zhou, Y., Xu, Ziying, Li, Z., Liu, Y., Ye, X., Li, X., Yuan, B., Ke, Y., He, C., Zhou, L., Liu, J. and Ci, W., 2016. Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer. *Cell Research*, 26(1), pp.103–118.

Chen, M.L., Shen, F., Huang, W., Qi, J.H., Wang, Y., Feng, Y.Q., Liu, S.M. and Yuan, B.F., 2013. Quantification of 5-methylcytosine and 5-hydroxymethylcytosine in

genomic DNA from hepatocellular carcinoma tissues by capillary hydrophilic-interaction liquid chromatography/quadrupole TOF mass spectrometry. *Clinical Chemistry*, 59(5), pp.824–832.

Chen, Q., Chen, Y., Bian, C., Fujiki, R. and Yu, X., 2013. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature*, 493(7433), pp.561–564.

Chen, Q., Espey, M.G., Sun, A.Y., Pooput, C., Kirk, K.L., Krishna, M.C., Khosh, D.B., Drisko, J. and Levine, M., 2008. Pharmacologic doses of ascorbate act as a prooxidant and decrease growth of aggressive tumor xenografts in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 105(32), pp.11105–11109.

Chen, Q., Yin, D., Zhang, Y., Yu, L., Li, X.D., Zhou, Z.J., Zhou, S.L., Gao, D.M., Hu, J., Jin, C., Wang, Z., Shi, Y.H., Cao, Y., Fan, J., Dai, Z. and Zhou, J., 2017. MicroRNA-29a induces loss of 5-hydroxymethylcytosine and promotes metastasis of hepatocellular carcinoma through a TET-SOCS1-MMP9 signaling axis. *Cell death & disease*, 8(6), p.e2906.

Chen, R., Zhang, Q., Duan, X., York, P., Chen, G.D., Yin, P., Zhu, H., Xu, M., Chen, P., Wu, Q., Li, D., Samarut, J., Xu, G., Zhang, P., Cao, X., Li, J. and Wong, J., 2017. The 5-hydroxymethylcytosine (5hmC) reader UHRF2 is required for normal levels of 5hmc in mouse adult brain and spatial learning and memory. *Journal of Biological Chemistry*, 292(11), pp.4533–4543.

Chen, S., Liu, Y., Wang, Y. and Xue, Z., 2019. LncRNA CCAT1 promotes colorectal cancer tumorigenesis via A miR-181b-5p/TUSC3 axis. *OncoTargets and Therapy*, 12, pp.9215–9225.

Chen, Z., Shi, X., Guo, L., Li, Y., Luo, M. and He, J., 2017. Decreased 5-hydroxymethylcytosine levels correlate with cancer progression and poor survival: A systematic review and meta-analysis. *Oncotarget*, 8(1), pp.1944–1952. Impact Journals LLC.

Cheng, J., Guo, S., Chen, S., Mastriano, S.J., Liu, C., D'Alessio, A.C., Hysolli, E., Guo, Y., Yao, H., Megyola, C.M., Li, D., Liu, J., Pan, W., Roden, C.A., Zhou, X.L., Heydari, K., Chen, J., Park, I.H., Ding, Y., Zhang, Y. and Lu, J., 2013. An Extensive Network of TET2-Targeting MicroRNAs Regulates Malignant

Hematopoiesis. *Cell Reports*, 5(2), pp.471–481.

Cheng, T.L., Chen, J., Wan, H., Tang, B., Tian, W., Liao, L. and Qiu, Z., 2017. Regulation of mRNA splicing by MeCP2 via epigenetic modifications in the brain. *Scientific Reports*, 7(1), pp.1–12.

Cheng, Y., He, C., Wang, M., Ma, X., Mo, F., Yang, S., Han, J. and Wei, X., 2019. Targeting epigenetic regulators for cancer therapy: Mechanisms and advances in clinical trials. *Signal Transduction and Targeted Therapy*, 4(1), pp.1–39.

Chmelarova, M., Krepinska, E., Spacek, J., Laco, J., Beranek, M. and Palicka, V., 2013. Methylation in the p53 promoter in epithelial ovarian cancer. *Clinical and Translational Oncology*, 15(2), pp.160–163.

Cho, Sungrae, Chae, J.S., Shin, H., Shin, Y., Song, H., Kim, Y., Yoo, B.C., Roh, K., Cho, Seungchan, Kil, E. joon, Byun, H. seong, Cho, Sang ho, Park, S., Lee, S. and Yeom, C.H., 2018. Hormetic dose response to L-ascorbic acid as an anti-cancer drug in colorectal cancer cell lines according to SVCT-2 expression. *Scientific Reports*, 8(1).

Choi, H.H., Cho, Y.-S., Choi, J.H., Kim, H.-K., Kim, S.S. and Chae, H.-S., 2019. Stool-Based miR-92a and miR-144* as Noninvasive Biomarkers for Colorectal Cancer Screening. *Oncology*, 97(3), pp.173–179.

Chomyk, A.M., Volsko, C., Tripathi, A., Deckard, S.A., Trapp, B.D., Fox, R.J. and Dutta, R., 2017. DNA methylation in demyelinated multiple sclerosis hippocampus. *Scientific Reports*, 7(1), pp.1–10.

Chou, C.-W., Wu, M.-S., Huang, W.-C. and Chen, C.-C., 2011. HDAC Inhibition Decreases the Expression of EGFR in Colorectal Cancer Cells. *PLoS ONE*, 6(3), p.e18087.

Chou, N.-H., Tsai, C.-Y., Tu, Y.-T., Wang, K.-C., Kang, C.-H., Chang, P.-M., Li, G.-C., Lam, H.-C., Liu, S.-I. and Tsai, K.-W., 2016. Isocitrate Dehydrogenase 2 Dysfunction Contributes to 5-hydroxymethylcytosine Depletion in Gastric Cancer Cells. *Anticancer research*, 36(8), pp.3983–90.

Choudhury, S.R., Cui, Y., Lubecka, K., Stefanska, B. and Irudayaraj, J., 2016. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1

promoter. *Oncotarget*, 7(29), pp.46545–46556.

Chowdhury, B., Cho, I.H. and Irudayaraj, J., 2017. Technical advances in global DNA methylation analysis in human cancers. *Journal of Biological Engineering*, 11(1), p.10. BioMed Central Ltd.

Chowdhury, R., Yeoh, K.K., Tian, Y., Hillringhaus, L., Bagg, E.A., Rose, N.R., Leung, I.K.H., Li, X.S., Woon, E.C.Y., Yang, M., McDonough, M.A., King, O.N., Clifton, I.J., Klose, R.J., Claridge, T.D.W., Ratcliffe, P.J., Schofield, C.J. and Kawamura, A., 2011. The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO reports*, 12(5), pp.463–469.

Choy, M.K., Movassagh, M., Goh, H.G., Bennett, M.R., Down, T.A. and Foo, R.S.Y., 2010. Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated. *BMC Genomics*, 11(1), pp.1–10.

Christiansen, L., Lenart, A., Tan, Q., Vaupel, J.W., Aviv, A., Mcgue, M. and Christensen, K., 2016. DNA methylation age is associated with mortality in a longitudinal Danish twin study. *Aging Cell*, 15(1), pp.149–154.

Chuang, K.H., Whitney-Miller, C.L., Chu, C.Y., Zhou, Z., Dokus, M.K., Schmit, S. and Barry, C.T., 2015. MicroRNA-494 is a master epigenetic regulator of multiple invasion-suppressor microRNAs by targeting ten eleven translocation 1 in invasive human hepatocellular carcinoma tumors. *Hepatology*, 62(2), pp.466–480.

Ciccarone, F., Valentini, E., Bacalini, M.G., Zampieri, M., Calabrese, R., Guastafierro, T., Mariano, G., Reale, A., Franceschi, C. and Caiafa, P., 2014. Poly(ADP-ribosyl)ation is involved in the epigenetic control of TET1 gene transcription. *Oncotarget*, 5(21), pp.10356–10367.

Ciccarone, F., Valentini, E., Zampieri, M. and Caiafa, P., 2015. 5mC-hydroxylase activity is influenced by the PARylation of TET1 enzyme. *Oncotarget*, 6(27), pp.24333–24347.

Ciechomska, M., Roszkowski, L. and Maslinski, W., 2019. DNA Methylation as a Future Therapeutic and Diagnostic Target in Rheumatoid Arthritis. *Cells*, 8(9), p.953.

Cimmino, L., Dawlaty, M.M., Ndiaye-Lobry, D., Yap, Y.S., Bakogianni, S., Yu, Y.,

Bhattacharyya, S., Shaknovich, R., Geng, H., Lobry, C., Mullenders, J., King, B., Trimarchi, T., Aranda-Orgilles, B., Liu, C., Shen, S., Verma, A.K., Jaenisch, R. and Aifantis, I., 2015. TET1 is a tumor suppressor of hematopoietic malignancy. *Nature Immunology*, 16(6), pp.653–662.

Cimmino, L., Neel, B.G. and Aifantis, I., 2018. Vitamin C in Stem Cell Reprogramming and Cancer. *Trends in Cell Biology*, 28(9), pp.698–708. Elsevier Ltd.

Claessens, H.A., 2001. Trends and progress in the characterization of stationary phases for reversed-phase liquid chromatography. *TrAC - Trends in Analytical Chemistry*, 20(10), pp.563–583.

Clark, A.E., Kaleta, E.J., Arora, A. and Wolk, D.M., 2013. Matrix-Assisted laser desorption ionization-time of flight mass spectrometry: A fundamental shift in the routine practice of clinical microbiology. *Clinical Microbiology Reviews*, 26(3), pp.547–603.

Coluccio, A., Ecco, G., Duc, J., Offner, S., Turelli, P. and Trono, D., 2018. Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naïve embryonic stem cells. *Epigenetics and Chromatin*, 11(1), pp.1–18.

Corbett, J.L. and Duncan, S.A., 2019. iPSC-Derived Hepatocytes as a Platform for Disease Modeling and Drug Discovery. *Frontiers in Medicine*, 6, p.265. Frontiers Media S.A.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P. V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., Das, S., Levasseur, D.N., Li, Z., Xu, M., Reik, W., Silva, J.C.R. and Wang, J., 2013. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature*, 495(7441), pp.370–374.

Creagan, E.T., Moertel, C.G., O'fallon, J.R., Schutt, A.J., O'connell, M.J., Rubin, J. and Frytak, S., 1979. Failure of High-Dose Vitamin C (Ascorbic Acid) Therapy to Benefit Patients with Advanced Cancer: A Controlled Trial. *New England Journal of Medicine*, 301(13), pp.687–690.

Creamer, K.M. and Lawrence, J.B., 2017. XIST RNA: a window into the broader role

of RNA in nuclear chromosome architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1733), p.20160360.

Crespo, M., Vilar, E., Tsai, S.Y., Chang, K., Amin, S., Srinivasan, T., Zhang, T., Pipalia, N.H., Chen, H.J., Witherspoon, M., Gordillo, M., Xiang, J.Z., Maxfield, F.R., Lipkin, S., Evans, T. and Chen, S., 2017. Colonic organoids derived from human induced pluripotent stem cells for modeling colorectal cancer and drug testing. *Nature Medicine*, 23(7), pp.878–884.

Cross, S.H., Meehan, R.R., Nan, X. and Bird, A., 1997. A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nature Genetics*, 16(3), pp.256–257.

Cui, H., Onyango, P., Brandenburg, S., Wu, Y., Hsieh, C.L. and Feinberg, A.P., 2002. Loss of imprinting in colorectal cancer linked to hypomethylation of H19 and IGF2. *Cancer Research*, 62(22), pp.6442–6446.

Cui, Q., Yang, S., Ye, P., Tian, E., Sun, Guoqiang, Zhou, J., Sun, Guihua, Liu, X., Chen, C., Murai, K., Zhao, C., Azizian, K.T., Yang, L., Warden, C., Wu, X., D'Apuzzo, M., Brown, C., Badie, B., Peng, L., Riggs, A.D., Rossi, J.J. and Shi, Y., 2016. Downregulation of TLX induces TET3 expression and inhibits glioblastoma stem cell self-renewal and tumorigenesis. *Nature Communications*, 7(1), pp.1–15.

Cui, Y., Li, T., Yang, D., Li, S. and Le, W., 2016. miR-29 regulates Tet1 expression and contributes to early differentiation of mouse ESCs. *Oncotarget*, 7(40), pp.64932–64941.

Cypris, O., Frobel, J., Rai, S., Franzen, J., Sontag, S., Goetzke, R., Szymanski De Toledo, M.A., Zenke, M. and Wagner, W., 2019. Tracking of epigenetic changes during hematopoietic differentiation of induced pluripotent stem cells. *Clinical Epigenetics*, 11(1), pp.1–11.

Dai, H., Hou, K., Cai, Z., Zhou, Q. and Zhu, S., 2017. Low-level miR-646 in colorectal cancer inhibits cell proliferation and migration by targeting NOB1 expression. *Oncology Letters*, 14(6), pp.6708–6714.

Dai, H.Q., Wang, B.A., Yang, L., Chen, J.J., Zhu, G.C., Sun, M.L., Ge, H., Wang, R., Chapman, D.L., Tang, F., Sun, X. and Xu, G.L., 2016. TET-mediated DNA

demethylation controls gastrulation by regulating Lefty-Nodal signalling. *Nature*, 538(7626), pp.528–532.

Dalerba, P., Dylla, S.J., Park, I.K., Liu, R., Wang, X., Cho, R.W., Hoey, T., Gurney, A., Huang, E.H., Simeone, D.M., Shelton, A.A., Parmiani, G., Castelli, C. and Clarke, M.F., 2007. Phenotypic characterization of human colorectal cancer stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 104(24), pp.10158–10163.

Dang, L., Jin, S. and Su, S.M., 2010. IDH mutations in glioma and acute myeloid leukemia. *Trends in Molecular Medicine*, 16(9), pp.387–397. Elsevier Ltd.

Darwiche, N., 2020. Epigenetic mechanisms and the hallmarks of cancer: an intimate affair. *American journal of cancer research*, 10(7), pp.1954–1978.

Dawlaty, M.M., Breiling, A., Le, T., Barrasa, M.I., Raddatz, G., Gao, Q., Powell, B.E., Cheng, A.W., Faull, K.F., Lyko, F. and Jaenisch, R., 2014. Loss of tet enzymes compromises proper differentiation of embryonic stem cells. *Developmental Cell*, 29(1), pp.102–111.

Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M., Faull, K.F., Lyko, F. and Jaenisch, R., 2013. Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development. *Developmental Cell*, 24(3), pp.310–323.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.W., Page, D.C. and Jaenisch, R., 2011. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell*, 9(2), pp.166–175.

Dekker, E., Tanis, P.J., Vleugels, J.L.A., Kasi, P.M. and Wallace, M.B., 2019. Colorectal cancer. *The Lancet*, 394(10207), pp.1467–1480. Lancet Publishing Group.

Deng, M., Zhang, R., He, Zhengxi, Qiu, Q., Lu, X., Yin, J., Liu, H., Jia, X. and He, Zhimin, 2017. TET-mediated sequestration of miR-26 drives EZH2 expression and gastric carcinogenesis. *Cancer Research*, 77(22), pp.6069–6082.

Deng, W., Wang, J., Zhang, J., Cai, J., Bai, Z. and Zhang, Z., 2016. TET2 regulates

LncRNA-ANRIL expression and inhibits the growth of human gastric cancer cells. *IUBMB Life*, 68(5), pp.355–364.

Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Méndez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., Shih, A.H., Levine, R.L., Bernard, O., Mercher, T., Solary, E., Urh, M., Daniels, D.L. and Fuks, F., 2013. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *The EMBO Journal*, 32(5), pp.645–655.

Ding, T., Cui, P., Zhou, Y., Chen, C., Zhao, J., Wang, H., Guo, M., He, Z. and Xu, L., 2018. Antisense Oligonucleotides against miR-21 Inhibit the Growth and Metastasis of Colorectal Carcinoma via the DUSP8 Pathway. *Molecular Therapy - Nucleic Acids*, 13, pp.244–255.

Doss, M.X. and Sachinidis, A., 2019. Current Challenges of iPSC-Based Disease Modeling and Therapeutic Implications. *Cells*, 8(5), p.403.

Dueland, S., Syversveen, T., Solheim, J.M., Solberg, S., Grut, H., Bjørnbeth, B.A., Hagness, M. and Line, P.-D., 2020. Survival Following Liver Transplantation for Patients With Nonresectable Liver-only Colorectal Metastases. *Annals of Surgery*, 271(2), pp.212–218.

Edwards, J.R., Yarychkivska, O., Boulard, M. and Bestor, T.H., 2017. DNA methylation and DNA methyltransferases. *Epigenetics and Chromatin*, 10(1), p.23. BioMed Central Ltd.

Eleftheriou, M., Pascual, A.J., Wheldon, L.M., Perry, C., Abakir, A., Arora, A., Johnson, A.D., Auer, D.T., Ellis, I.O., Madhusudan, S. and Ruzov, A., 2015. 5-Carboxylcytosine levels are elevated in human breast cancers and gliomas. *Clinical Epigenetics*, 7(1), p.88.

Van Emburgh, B.O. and Robertson, K.D., 2011. Modulation of Dnmt3b function in vitro by interactions with Dnmt3L, Dnmt3a and Dnmt3b splice variants. *Nucleic Acids Research*, 39(12), pp.4984–5002.

Engstrand, J., Nilsson, H., Strömberg, C., Jonas, E. and Freedman, J., 2018. Colorectal cancer liver metastases – a population-based study on incidence, management and survival. *BMC Cancer*, 18(1), p.78.

Erkens, T., Van Poucke, M., Vandesompele, J., Goossens, K., Van Zeveren, A. and Peelman, L.J., 2006. Development of a new set of reference genes for normalization of real-time RT-PCR data of porcine backfat and longissimus dorsi muscle, and evaluation with PPARGC1A. *BMC Biotechnology*, 6(1), pp.1–8.

Esteban, M.A., Wang, T., Qin, B., Yang, J., Qin, D., Cai, J., Li, W., Weng, Z., Chen, J., Ni, S., Chen, K., Li, Y., Liu, X., Xu, J., Zhang, S., Li, F., He, W., Labuda, K., Song, Y., Peterbauer, A., Wolbank, S., Redl, H., Zhong, M., Cai, D., Zeng, L. and Pei, D., 2010. Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells. *Cell Stem Cell*, 6(1), pp.71–79.

Fain, O., Mathieu, E. and Thomas, M., 1998. Lesson of the week. Scurvy in patients with cancer. *British Medical Journal*, 316(7145), pp.1661–1662. BMJ Publishing Group.

Farthing, C.R., Ficz, G., Ng, R.K., Chan, C.-F., Andrews, S., Dean, W., Hemberger, M. and Reik, W., 2008. Global Mapping of DNA Methylation in Mouse Promoters Reveals Epigenetic Reprogramming of Pluripotency Genes. *PLoS Genetics*, 4(6), p.e1000116.

Fearon, E.F. and Vogelstein, B., 1990. *A Genetic Model for Colorectal Tumorigenesis*.

Feinberg, A.P. and Vogelstein, B., 1983a. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, 301(5895), pp.89–92.

Feinberg, A.P. and Vogelstein, B., 1983b. Hypomethylation of ras oncogenes in primary human cancers. *Biochemical and Biophysical Research Communications*, 111(1), pp.47–54.

Feng, J., Wang, Q., Li, G., Zeng, X., Kuang, S., Li, X. and Yue, Y., 2015. TET1-mediated different transcriptional regulation in prostate cancer. *International Journal of Clinical and Experimental Medicine*, 8(1), pp.203–211.

Fernandez, A.F., Bayón, G.F., Sierra, M.I., Urdinguio, R.G., Toraño, E.G., García, M.G., Carella, A., López, V., Santamarina, P., Pérez, R.F., Belmonte, T., Tejedor, J.R., Cobo, I., Menendez, P., Mangas, C., Ferrero, C., Rodrigo, L., Astudillo, A., Ortea, I., Díaz, S.C., Rodríguez-Gonzalez, P., Alonso, J.I.G., Mollejo, M., Meléndez, B., Domínguez, G., Bonilla, F. and Fraga, M.F., 2018. Loss of 5hmC

identifies a new type of aberrant DNA hypermethylation in glioma. *Human Molecular Genetics*, 27(17), pp.3046–3059.

Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J., Cook, S.J., Andrews, S., Branco, M.R. and Reik, W., 2013. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell*, 13(3), pp.351–359.

Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., Tallman, M.S., Sun, Z., Wolniak, K., Peeters, J.K., Liu, W., Choe, S.E., Fantin, V.R., Paietta, E., Löwenberg, B., Licht, J.D., Godley, L.A., Delwel, R., Valk, P.J.M., Thompson, C.B., Levine, R.L. and Melnick, A., 2010. Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell*, 18(6), pp.553–567.

Filipczak, P.T., Leng, S., Tellez, C.S., Do, K.C., Grimes, M.J., Thomas, C.L., Walton-Filipczak, S.R., Picchi, M.A. and Belinsky, S.A., 2019. P53-suppressed oncogene TET1 prevents cellular aging in lung cancer. *Cancer Research*, 79(8), pp.1758–1768.

Flavahan, W.A., Gaskell, E. and Bernstein, B.E., 2017. Epigenetic plasticity and the hallmarks of cancer. *Science*, 357(6348). American Association for the Advancement of Science.

Forn, M., Díez-Villanueva, A., Merlos-Suárez, A., Muñoz, M., Lois, S., Carriò, E., Jordà, M., Bigas, A., Batlle, E. and Peinado, M.A., 2015. Overlapping DNA methylation dynamics in mouse intestinal cell differentiation and early stages of malignant progression. *PLoS ONE*, 10(5).

Fu, F., Jiang, W., Zhou, L. and Chen, Z., 2018. Circulating Exosomal miR-17-5p and miR-92a-3p Predict Pathologic Stage and Grade of Colorectal Cancer. *Translational Oncology*, 11(2), pp.221–232.

Fu, T., Liu, L., Yang, Q.L., Wang, Y., Xu, P., Zhang, Lin, Liu, S., Dai, Q., Ji, Q., Xu, G.L., He, C., Luo, C. and Zhang, Liang, 2019. Thymine DNA glycosylase recognizes the geometry alteration of minor grooves induced by 5-formylcytosine and 5-carboxylcytosine. *Chemical Science*, 10(31), pp.7407–7417.

Fu, X., Jin, L., Wang, X., Luo, A., Hu, J., Zheng, X., Tsark, W.M., Riggs, A.D., Ku, H.T. and Huang, W., 2013. MicroRNA-26a targets ten eleven translocation enzymes and is regulated during pancreatic cell differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(44), pp.17892–17897.

Fumagalli, A., Oost, K.C., Kester, L., Morgner, J., Bornes, L., Bruens, L., Spaargaren, L., Azkanaz, M., Schelfhorst, T., Beerling, E., Heinz, M.C., Postrach, D., Seinstra, D., Sieuwerts, A.M., Martens, J.W.M., van der Elst, S., van Baalen, M., Bhowmick, D., Vrisekoop, N., Ellenbroek, S.I.J., Suijkerbuijk, S.J.E., Snippert, H.J. and van Rheenen, J., 2020. Plasticity of Lgr5-Negative Cancer Cells Drives Metastasis in Colorectal Cancer. *Cell Stem Cell*, 26(4), pp.569-578.e7.

Gailhouste, L., Liew, L.C., Yasukawa, K., Hatada, I., Tanaka, Y., Kato, T., Nakagama, H. and Ochiya, T., 2019. MEG3-derived miR-493-5p overcomes the oncogenic feature of IGF2-miR-483 loss of imprinting in hepatic cancer cells. *Cell Death and Disease*, 10(8), pp.1–16.

Gama-sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W. and Ehrlich, M., 1983. The 5-methylcytosine content of DNA from human tumors. *Nucleic Acids Research*, 11(19), pp.6883–6894.

Gan, L., Camarena, V., Mustafi, S. and Wang, G., 2019. Vitamin C inhibits triple-negative breast cancer metastasis by affecting the expression of YAP1 and synaptopodin 2. *Nutrients*, 11(12).

Gao, P., Lin, S., Cai, M., Zhu, Y., Song, Y., Sui, Y., Lin, J., Liu, J., Lu, X., Zhong, Y., Cui, Y. and Zhou, P., 2019. 5-Hydroxymethylcytosine profiling from genomic and cell-free DNA for colorectal cancers patients. *Journal of Cellular and Molecular Medicine*, 23(5), pp.3530–3537.

Gao, Yawei, Chen, Jiayu, Li, K., Wu, T., Huang, B., Liu, W., Kou, X., Zhang, Y., Huang, H., Jiang, Y., Yao, C., Liu, X., Lu, Z., Xu, Z., Kang, L., Chen, Jun, Wang, H., Cai, T. and Gao, S., 2013. Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell*, 12(4), pp.453–469.

Gao, Yuan, Yang, L., Chen, L., Wang, X., Wu, H., Ai, Z., Du, J., Liu, Y., Shi, X., Wu,

Y., Guo, Z. and Zhang, Y., 2013. Vitamin C facilitates pluripotent stem cell maintenance by promoting pluripotency gene transcription. *Biochimie*, 95(11), pp.2107–2113.

García, M.G., Carella, A., Urdinguio, R.G., Bayón, G.F., Lopez, V., Tejedor, J.R., Sierra, M.I., García-Toraño, E., Santamarina, P., Perez, R.F., Mangas, C., Astudillo, A., Corte-Torres, M.D., Sáenz-de-Santa-María, I., Chiara, M.D., Fernández, A.F. and Fraga, M.F., 2018. Epigenetic dysregulation of TET2 in human glioblastoma. *Oncotarget*, 9(40), pp.25922–25934.

Ge, G., Peng, D., Xu, Z., Guan, B., Xin, Z., He, Q., Zhou, Y., Li, X., Zhou, L. and Ci, W., 2018. Restoration of 5-hydroxymethylcytosine by ascorbate blocks kidney tumour growth. *EMBO reports*, 19(8).

Gerecke, C., Schumacher, F., Berndzen, A., Homann, T. and Kleuser, B., 2020. Vitamin C in combination with inhibition of mutant IDH1 synergistically activates TET enzymes and epigenetically modulates gene silencing in colon cancer cells. *Epigenetics*, 15(3), pp.307–322.

Gibbons, R.J., McDowell, T.L., Raman, S., O'Rourke, D.M., Garrick, D., Ayyub, H. and Higgs, D.R., 2000. Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nature Genetics*, 24(4), pp.368–371.

Gilat, N., Tabachnik, T., Shwartz, A., Shahal, T., Torchinsky, D., Michaeli, Y., Nifker, G., Zirkin, S. and Ebenstein, Y., 2017. Single-molecule quantification of 5-hydroxymethylcytosine for diagnosis of blood and colon cancers. *Clinical Epigenetics*, 9(1), pp.1–8.

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M. and Carell, T., 2010. Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. *PLoS ONE*, 5(12), p.e15367.

Glowacka, W.K., Jain, H., Okura, M., Maimaitiming, A., Mamatjan, Y., Nejad, R., Farooq, H., Taylor, M.D., Aldape, K. and Kongkham, P., 2018. 5-Hydroxymethylcytosine preferentially targets genes upregulated in isocitrate dehydrogenase 1 mutant high-grade glioma. *Acta Neuropathologica*, 135(4), pp.617–634.

Goldstone, A.P., 2004. Prader-Willi syndrome: Advances in genetics, pathophysiology and treatment. *Trends in Endocrinology and Metabolism*, 15(1), pp.12–20. Elsevier Current Trends.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E. and Bestor, T.H., 2006. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science*, 311(5759), pp.395–398.

Good, C.R., Madzo, J., Patel, B., Maegawa, S., Engel, N., Jelinek, J. and Issa, J.P.J., 2017. A novel isoform of TET1 that lacks a CXXC domain is overexpressed in cancer. *Nucleic Acids Research*, 45(14), pp.8269–8281.

Good, C.R., Panjarian, S., Kelly, A.D., Madzo, J., Patel, B., Jelinek, J. and Issa, J.P.J., 2018. TET1-Mediated hypomethylation activates oncogenic signaling in triple-Negative breast cancer. *Cancer Research*, 78(15), pp.4126–4137.

Gorgen, A., Muaddi, H., Zhang, W., McGilvray, I., Gallinger, S. and Sapisochin, G., 2018. The New Era of Transplant Oncology: Liver Transplantation for Nonresectable Colorectal Cancer Liver Metastases. *Canadian Journal of Gastroenterology and Hepatology*, 2018. Hindawi Limited.

Gowher, H., Liebert, K., Hermann, A., Xu, G. and Jeltsch, A., 2005. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *Journal of Biological Chemistry*, 280(14), pp.13341–13348.

Greger, V., Debus, N., Lohmann, D., Höpping, W., Passarge, E. and Horsthemke, B., 1994. Frequency and parental origin of hypermethylated RB1 alleles in retinoblastoma. *Human Genetics*, 94(5), pp.491–496.

Greger, V., Passarge, E., Hfpping, W. and Horsthemke, B., 1989. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Human Genetics*, 83, pp.155–158.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.L.G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G., Iqbal, K., Shi, Y.G., Deng, Z., Szabó, P.E., Pfeifer, G.P., Li, J. and Xu, G.L.G.F., 2011. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, 477(7366), pp.606–612.

Guibert, S., Forné, T. and Weber, M., 2012. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Research*, 22(4), pp.633–641.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.P., Hu, B., Walsh, C.P., Li, J., Tang, F. and Xu, G.L., 2014. Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell*, 15(4), pp.447–459.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., Jin, X., Shi, X., Liu, P., Wang, X., Wang, W., Wei, Y., Li, X., Guo, F., Wu, X., Fan, X., Yong, J., Wen, L., Xie, S.X., Tang, F. and Qiao, J., 2014. The DNA methylation landscape of human early embryos. *Nature*, 511(7511), pp.606–610.

Guo, M., Li, X., Zhang, L., Liu, D., Du, W., Yin, D., Lyu, N., Zhao, G., Guo, C. and Tang, D., 2017. Accurate quantification of 5-Methylcytosine, 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine in genomic DNA from breast cancer by chemical derivatization coupled with ultra performance liquid chromatography- electrospray quadrupole time of flight mass spectrometry analysis. *Oncotarget*, 8(53), pp.91248–91257.

Gupta, P.B., Onder, T.T., Jiang, G., Tao, K., Kuperwasser, C., Weinberg, R.A. and Lander, E.S., 2009. Identification of Selective Inhibitors of Cancer Stem Cells by High-Throughput Screening. *Cell*, 138(4), pp.645–659.

Gustafson, C.B., Yang, C., Dickson, K.M., Shao, H., Van Booven, D., Harbour, J.W., Liu, Z.J. and Wang, G., 2015. Epigenetic reprogramming of melanoma cells by vitamin C treatment. *Clinical Epigenetics*, 7(1).

Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G. and Surani, M.A., 2013. Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency. *Stem Cell Reports*, 1(6), pp.518–531.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A. and Surani, M.A., 2013. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science*, 339(6118), pp.448–452.

Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T., Yoshida, K., Roller, A.,

Nadarajah, N., Shiraishi, Y., Shiozawa, Y., Chiba, K., Tanaka, H., Koeffler, H.P., Klein, H.U., Dugas, M., Aburatani, H., Kohlmann, A., Miyano, S., Haferlach, C., Kern, W. and Ogawa, S., 2014. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*, 28(2), pp.241–247.

Haffner, M.C., Chaux, A., Meeker, A.K., Esopi, D.M., Gerber, J., Pellakuru, L.G., Toubaji, A., Argani, P., Iacobuzio-Donahue, C., Nelson, W.G., Netto, G.J., De Marzo, A.M. and Yegnasubramanian, S., 2011. Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget*, 2(8), pp.627–637.

Hanahan, D. and Weinberg, R.A., 2000. The hallmarks of cancer. *Cell*, 100(1), pp.57–70. Elsevier.

Hanahan, D. and Weinberg, R.A., 2011. Hallmarks of cancer: The next generation. *Cell*, 144(5), pp.646–674. Cell Press.

Hanai, T., 2017. Quantitative Explanation of Retention Mechanisms of Hydrophobic and Hydrophilic-Interaction Liquid Chromatography-Inductive Effect of Alkyl Chain. *Separations*, 4(4), p.33.

Hanna, C.W. and Kelsey, G., 2014. The specification of imprints in mammals. *Heredity*, 113(2), pp.176–183. Nature Publishing Group.

Hardwick, J.S., Ptchelkine, D., El-Sagheer, A.H., Tear, I., Singleton, D., Phillips, S.E.V., Lane, A.N. and Brown, T., 2017. 5-Formylcytosine does not change the global structure of DNA. *Nature Structural and Molecular Biology*, 24(6), pp.544–552.

Hargan-Calvopina, J., Taylor, S., Cook, H., Hu, Z., Lee, S.A., Yen, M.R., Chiang, Y.S., Chen, P.Y. and Clark, A.T., 2016. Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. *Developmental Cell*, 39(1), pp.75–86.

Haridhasapavalan, K.K., Raina, K., Dey, C., Adhikari, P. and Thummer, R.P., 2020. An Insight into Reprogramming Barriers to iPSC Generation. *Stem Cell Reviews and Reports*, 16(1), pp.56–81. Springer.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M.,

Zhang, X. and Cheng, X., 2012. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research*, 40(11), pp.4841–4849.

Hata, K., Okano, M., Lei, H. and Li, E., 2002. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*, 129(8), pp.1983–1993.

Hay, D.C., Zhao, D., Fletcher, J., Hewitt, Z.A., McLean, D., Urruticoechea-Uriguen, A., Black, J.R., Elcombe, C., Ross, J.A., Wolf, R. and Cui, W., 2008. Efficient Differentiation of Hepatocytes from Human Embryonic Stem Cells Exhibiting Markers Recapitulating Liver Development In Vivo. *Stem Cells*, 26(4), pp.894–902.

He, S., Sun, H., Lin, L., Zhang, Y., Chen, Jinlong, Liang, L., Li, Y., Zhang, M., Yang, X., Wang, X., Wang, F., Zhu, F., Chen, Jiekai, Pei, D. and Zheng, H., 2017. Passive DNA demethylation preferentially up-regulates pluripotency-related genes and facilitates the generation of induced pluripotent stem cells. *Journal of Biological Chemistry*, 292(45), pp.18542–18555.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, N., Sun, Y., Li, X., Dai, Q., Song, C.X., Zhang, K., He, C. and Xu, G.L., 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, 333(6047), pp.1303–1307.

Hedrich, C.M., Mäbert, K., Rauen, T. and Tsokos, G.C., 2017. DNA methylation in systemic lupus erythematosus. *Epigenomics*, 9(4), pp.505–525. Future Medicine Ltd.

Van Der Heijden, M. and Vermeulen, L., 2019. Stem cells in homeostasis and cancer of the gut. *Molecular Cancer*, 18(1), p.66. BioMed Central Ltd.

Hershey, A.D., Dixon, J. and Chase, M., 1953. Nucleic acid economy in bacteria infected with bacteriophage T2. I. Purine and pyrimidine composition. *The Journal of general physiology*, 36(6), pp.777–789.

Hewitt, R.E., McMarlin, A., Kleiner, D., Wersto, R., Martin, P., Tsoskas, M., Stamp, G.W.H. and Stetler-Stevenson, W.G., 2000. Validation of a model of colon cancer

progression. *The Journal of Pathology*, 192(4), pp.446–454.

Hibi, K., Nakamura, H., Hirai, A., Fujikake, Y., Kasai, Y., Akiyama, S., Ito, K. and Takagi, H., 1996. Loss of H19 Imprinting in Esophageal Cancer. *Cancer Research*, 56(3).

Hildebrandt, M.R., Germain, D.R., Monckton, E.A., Brun, M. and Godbout, R., 2015. Ddx1 knockout results in transgenerational wild-type lethality in mice. *Scientific Reports*, 5(1), pp.1–8.

Hill, P.W.S., Leitch, H.G., Requena, C.E., Sun, Z., Amouroux, R., Roman-Trufero, M., Borkowska, M., Terragni, J., Vaisvila, R., Linnett, S., Bagci, H., Dharmalingham, G., Haberle, V., Lenhard, B., Zheng, Y., Pradhan, S. and Hajkova, P., 2018. Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature*, 555(7696), pp.392–396.

Hiramatsu, Y., Fukuda, A., Ogawa, S., Goto, N., Ikuta, K., Tsuda, M., Matsumoto, Y., Kimura, Y., Yoshioka, T., Takada, Y., Maruno, T., Hanyu, Y., Tsuruyama, T., Wang, Z., Akiyama, H., Takaishi, S., Miyoshi, H., Taketo, M.M., Chiba, T. and Seno, H., 2019. Arid1a is essential for intestinal stem cells through Sox9 regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 116(5), pp.1704–1713.

Hirano, K.I., Young, S.G., Farese, R. V., Ng, J., Sande, E., Warburton, C., Powell-Braxton, L.M. and Davidson, N.O., 1996. Targeted disruption of the mouse apobec-1 gene abolishes apolipoprotein B mRNA editing and eliminates apolipoprotein B48. *Journal of Biological Chemistry*, 271(17), pp.9887–9890.

Hlady, R.A., Sathyanarayan, A., Thompson, J.J., Zhou, D., Wu, Q., Pham, K., Lee, J., Liu, C. and Robertson, K.D., 2019. Integrating the Epigenome to Identify Drivers of Hepatocellular Carcinoma. *Hepatology*, 69(2), pp.639–652.

Hochberg, I., Harvey, I., Tran, Q.T., Stephenson, E.J., Barkan, A.L., Saltiel, A.R., Chandler, W.F. and Bridges, D., 2015. Gene expression changes in subcutaneous adipose tissue due to cushing's disease. *Journal of Molecular Endocrinology*, 55(2), pp.81–94.

Hofer, A., Liu, Z.J. and Balasubramanian, S., 2019. Detection, Structure and Function

of Modified DNA Bases. *Journal of the American Chemical Society*, 141(16), pp.6420–6429.

Hoffman, M.D., Sniatynski, M.J., Rogalski, J.C., Le Blanc, J.C.Y. and Kast, J., 2006. Multiple neutral loss monitoring (MNM): A multiplexed method for post-translational modification screening. *Journal of the American Society for Mass Spectrometry*, 17(3), pp.307–317.

Holliday, R. and Pugh, J.E., 1975. DNA Modification Mechanisms and Gene Activity during Development. *Science*, 187(4173), pp.226–232.

Holmes, D., 2015. A disease of growth. *Nature*, 521(7551), pp.S2–S3.

Holtzinger, A., Streeter, P.R., Sarangi, F., Hillborn, S., Niapour, M., Ogawa, S. and Keller, G., 2015. New markers for tracking endoderm induction and hepatocyte differentiation from human pluripotent stem cells. *Development (Cambridge)*, 142(24), pp.4253–4265.

Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D. and Ren, B., 2013. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature Genetics*, 45(10), pp.1198–1206.

Hon, G.C., Song, C.X., Du, T., Jin, F., Selvaraj, S., Lee, A.Y., Yen, C.A., Ye, Z., Mao, S.Q., Wang, B.A., Kuan, S., Edsall, L.E., Zhao, B.S., Xu, G.L., He, C. and Ren, B., 2014. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Molecular Cell*, 56(2), pp.286–297.

Hore, T.A., Von Meyenn, F., Ravichandran, M., Bachman, M., Ficz, G., Oxley, D., Santos, F., Balasubramanian, S., Jurkowski, T.P. and Reik, W., 2016. Retinol and ascorbate drive erasure of Epigenetic memory and enhance reprogramming to naïve pluripotency by complementary mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*, 113(43), pp.12202–12207.

Horvath, S., 2013. DNA methylation age of human tissues and cell types. *Genome Biology*, 14(10), pp.1–20.

Horvath, S., 2015. Erratum to DNA methylation age of human tissues and cell types [Genome Biology, 14, R115, (2013)]. *Genome Biology*, 16(1), pp.1–5.

Horvath, S. and Ritz, B.R., 2015. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging*, 7(12), pp.1130–1142.

Hotchkiss, R.D., 1948. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *Journal of Biological Chemistry*, 175, pp.315–332.

Hrit, J., Goodrich, L., Li, C., Wang, B.A., Nie, J., Cui, X., Martin, E.A., Simental, E., Fernandez, J., Liu, M.Y., Nery, J.R., Castanon, R., Kohli, R.M., Tretyakova, N., He, C., Ecker, J.R., Goll, M. and Panning, B., 2018. OGT binds a conserved C-terminal domain of TET1 to regulate TET1 activity and function in development. *eLife*, 7.

Hu, H., Shu, M., He, L., Yu, X., Liu, X., Lu, Y., Chen, Y., Miao, X. and Chen, X., 2017. Epigenomic landscape of 5-hydroxymethylcytosine reveals its transcriptional regulation of lncRNAs in colorectal cancer. *British Journal of Cancer*, 116(5), pp.658–668.

Hu, X., Zhang, L., Mao, S.Q., Li, Z., Chen, J., Zhang, R.R., Wu, H.P., Gao, J., Guo, F., Liu, W., Xu, G.F., Dai, H.Q., Shi, Y.G., Li, X., Hu, B., Tang, F., Pei, D. and Xu, G.L., 2014. Tet and TDG mediate DNA demethylation essential for mesenchymal-to- epithelial transition in somatic cell reprogramming. *Cell Stem Cell*, 14(4), pp.512–522.

Huang, D., Sun, W., Zhou, Y., Li, P., Chen, F., Chen, H., Xia, D., Xu, E., Lai, M., Wu, Y. and Zhang, H., 2018. Mutations of key driver genes in colorectal cancer progression and metastasis. *Cancer and Metastasis Reviews*, 37(1), pp.173–187. Springer New York LLC.

Huang, H., Jiang, X., Li, Z., Li, Y., Song, C.X., He, Chunjiang, Sun, M., Chen, P., Gurbuxani, S., Wang, Jiapeng, Hong, G.M., Elkahloun, A.G., Arnovitz, S., Wang, Jinhua, Szulwach, K., Lin, L., Street, C., Wunderlich, M., Dawlaty, M., Neilly, M.B., Jaenisch, R., Yang, F.C., Mulloy, J.C., Jin, P., Liu, P.P., Rowley, J.D., Xu, M., He, Chuan and Chen, J., 2013. TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, 110(29), pp.11994–11999.

Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W.A., Kang, J., Zepeda-Martínez,

J.A., Pape, U.J., Jacobsen, S.E., Peters, B. and Rao, A., 2014. Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 111(4), pp.1361–1366.

Huang, Y. and Rao, A., 2014. Connections between TET proteins and aberrant DNA modification in cancer. *Trends in Genetics*, 30(10), pp.464–474. Elsevier Ltd.

Huang, Y., Wang, G., Liang, Z., Yang, Y., Cui, L. and Liu, C.-Y., 2016. Loss of nuclear localization of TET2 in colorectal cancer. *Clinical Epigenetics*, 8(1), p.9.

Iancu, I. V., Botezatu, A., Plesa, A., Huica, I., Fudulu, A., Albulescu, A., Bostan, M., Mihaila, M., Grancea, C., Manda, D.A., Dobrescu, R., Vladoiu, S.V., Anton, G. and Badiu, C.V., 2020. Alterations of regulatory factors and DNA methylation pattern in thyroid cancer. *Cancer Biomarkers*, 28(2), pp.255–268.

Ibrahim, H., Dresser, K., Cornejo, K.M., Ibrahim, H., Dresser, K. and Cornejo, K.M., 2019. Loss of 5-hydroxymethylcytosine (5-hmC) expression in endometrioid type endometrial carcinoma. *European Journal of Gynaecological Oncology*, 40(5), pp.846–848.

Igder, S., Mohammadiasl, J. and Mokarram, P., 2019. Altered miR-21, miRNA-148a Expression in Relation to KRAS Mutation Status as Indicator of Adenoma-Carcinoma Transitional Pattern in Colorectal Adenoma and Carcinoma Lesions. *Biochemical Genetics*, 57(6), pp.767–780.

Inoue, S., Lemonnier, F. and Mak, T.W., 2016. Roles of IDH1/2 and TET2 mutations in myeloid disorders. *International Journal of Hematology*, 103(6), pp.627–633. Springer Tokyo.

Iqbal, K., Jin, S.G., Pfeifer, G.P. and Szabó, P.E., 2011. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proceedings of the National Academy of Sciences of the United States of America*, 108(9), pp.3642–3647.

Itakura, G., Kawabata, S., Ando, M., Nishiyama, Y., Sugai, K., Ozaki, M., Iida, T., Ookubo, T., Kojima, K., Kashiwagi, R., Yasutake, K., Nakauchi, H., Miyoshi, H., Nagoshi, N., Kohyama, J., Iwanami, A., Matsumoto, M., Nakamura, M. and

Okano, H., 2017. Fail-Safe System against Potential Tumorigenicity after Transplantation of iPSC Derivatives. *Stem Cell Reports*, 8(3), pp.673–684.

Ito, Kyoko, Lee, J., Chrysanthou, S., Zhao, Y., Josephs, K., Sato, H., Teruya-Feldstein, J., Zheng, D., Dawlaty, M.M. and Ito, Keisuke, 2019. Non-catalytic Roles of Tet2 Are Essential to Regulate Hematopoietic Stem and Progenitor Cell Homeostasis. *Cell Reports*, 28(10), pp.2480-2490.e4.

Ito, R., Katsura, S., Shimada, H., Tsuchiya, H., Hada, M., Okumura, T., Sugawara, A. and Yokoyama, A., 2014. TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes to Cells*, 19(1), pp.52–65.

Ito, S., Dalessio, A.C., Taranova, O. V., Hong, K., Sowers, L.C. and Zhang, Y., 2010. Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 466(7310), pp.1129–1133.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y., 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, 333(6047), pp.1300–1303.

Itsumi, M., Inoue, S., Elia, A.J., Murakami, K., Sasaki, M., Lind, E.F., Brenner, D., Harris, I.S., Chio, I.I.C., Afzal, S., Cairns, R.A., Cescon, D.W., Elford, A.R., Ye, J., Lang, P.A., Li, W.Y., Wakeham, A., Duncan, G.S., Haight, J., You-Ten, A., Snow, B., Yamamoto, K., Ohashi, P.S. and Mak, T.W., 2015. Idh1 protects murine hepatocytes from endotoxin-induced oxidative stress by regulating the intracellular NADP +/NADPH ratio. *Cell Death and Differentiation*, 22(11), pp.1837–1848.

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E.A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S. and Reik, W., 2013. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biology*, 14(10), p.R119.

Iurlaro, M., McInroy, G.R., Burgess, H.E., Dean, W., Raiber, E.A., Bachman, M., Beraldi, D., Balasubramanian, S. and Reik, W., 2016. In vivo genome-wide profiling reveals a tissue-specific role for 5-formylcytosine. *Genome Biology*, 17(1), p.141.

Ivanov, M., Kals, M., Kacevska, M., Barragan, I., Kasuga, K., Rane, A., Metspalu, A.,

Milani, L. and Ingelman-Sundberg, M., 2013. Ontogeny, distribution and potential roles of 5-hydroxymethylcytosine in human liver function. *Genome Biology*, 14(8), pp.1–15.

Jacob, F., Guertler, R., Naim, S., Nixdorf, S., Fedier, A., Hacker, N.F. and Heinzelmann-Schwarz, V., 2013. Careful Selection of Reference Genes Is Required for Reliable Performance of RT-qPCR in Human Normal and Cancer Cell Lines. *PLoS ONE*, 8(3).

Jang, H.S., Shin, W.J., Lee, J.E. and Do, J.T., 2017. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes*, 8(6), p.148.

Janke, R., Iavarone, A.T. and Rine, J., 2017. Oncometabolite d-2-hydroxyglutarate enhances gene silencing through inhibition of specific h3k36 histone demethylases. *eLife*, 6.

Jansen, A.M.L., Ghosh, P., Dakal, T.C., Slavin, T.P., Boland, C.R. and Goel, A., 2020. Novel candidates in early-onset familial colorectal cancer. *Familial Cancer*, 19(1), pp.1–10.

Jäwert, F., Hasséus, B., Kjeller, G., Magnusson, B., Sand, L. and Larsson, L., 2013. Loss of 5-hydroxymethylcytosine and TET2 in oral squamous cell carcinoma. *Anticancer Research*, 33(10), p.4328.

Jeong, J.J., Gu, X., Nie, J., Sundaravel, S., Liu, H., Kuo, W.L., Bhagat, T.D., Pradhan, K., Cao, J., Nischal, S., McGraw, K.L., Bhattacharyya, S., Bishop, M.R., Artz, A., Thirman, M.J., Moliterno, A., Ji, P., Levine, R.L., Godley, L.A., Steidl, U., Bieker, J.J., List, A.F., Saunthararajah, Y., He, C., Verma, A. and Wickrema, A., 2019. Cytokine-regulated phosphorylation and activation of TET2 by JAK2 in hematopoiesis. *Cancer Discovery*, 9(6), pp.778–795.

Ji, D., Lin, K., Song, J. and Wang, Y., 2014. Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. *Molecular BioSystems*, 10(7), pp.1749–1752.

Ji, S., Park, D., Kropachev, K., Kolbanovskiy, M., Fu, I., Broyde, S., Essawy, M., Geacintov, N.E. and Tretyakova, N.Y., 2019. 5-Formylcytosine-induced DNA–peptide cross-links reduce transcription efficiency, but do not cause transcription

errors in human cells. *Journal of Biological Chemistry*, 294(48), pp.18387–18397.

Jia, M., Gao, X., Zhang, Y., Hoffmeister, M. and Brenner, H., 2016. Different definitions of CpG island methylator phenotype and outcomes of colorectal cancer: a systematic review. *Clinical Epigenetics*, 8(1), pp.1–14. Springer Verlag.

Jia, M., Li, Z., Pan, M., Tao, M., Wang, J. and Lu, X., 2020. LINC-PINT suppresses aggressiveness of thyroid cancer by downregulating miR-767-5p to induce TET2 expression. *Molecular Therapy - Nucleic Acids*.

Jia, Z., Liang, Y., Ma, B., Xu, X., Xiong, J., Duan, L. and Wang, D., 2017. A 5-mC dot blot assay quantifying the DNA methylation level of chondrocyte dedifferentiation in vitro. *Journal of Visualized Experiments*, 2017(123).

Jiang, D., Wei, S., Chen, F., Zhang, Y. and Li, J., 2017. TET3-mediated DNA oxidation promotes ATR-dependent DNA damage response. *EMBO reports*, 18(5), pp.781–796.

Jiang, W., Liu, Y., Liu, R., Zhang, K. and Zhang, Y., 2015. The lncRNA DEANR1 facilitates human endoderm differentiation by activating FOXA2 expression. *Cell Reports*, 11(1), pp.137–148.

Jin, S.G., Jiang, Y., Qiu, R., Rauch, T.A., Wang, Y., Schackert, G., Krex, D., Lu, Q. and Pfeifer, G.P., 2011. 5-hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Research*, 71(24), pp.7360–7365.

Jin, S.G., Zhang, Z.M., Dunwell, T.L., Harter, M.R., Wu, X., Johnson, J., Li, Z., Liu, J., Szabó, P.E., Lu, Q., Xu, G. liang, Song, J. and Pfeifer, G.P., 2016. Tet3 Reads 5-Carboxylcytosine through Its CXXC Domain and Is a Potential Guardian against Neurodegeneration. *Cell Reports*, 14(3), pp.493–505.

Jögi, A., Vaapil, M., Johansson, M. and Påhlman, S., 2012. Cancer cell differentiation heterogeneity and aggressive behavior in solid tumors. *Upsala Journal of Medical Sciences*, 117(2), pp.217–224. Taylor & Francis.

Johnson, K.C., Houseman, E.A., King, J.E., Von Herrmann, K.M., Fadul, C.E. and Christensen, B.C., 2016. 5-Hydroxymethylcytosine localizes to enhancer elements and is associated with survival in glioblastoma patients. *Nature Communications*,

7(1), pp.1–11.

Johnson, M.H. and Cohen, J., 2012. Reprogramming rewarded: The 2012 nobel prize for physiology or medicine awarded to John Gurdon and Shinya Yamanaka. *Reproductive BioMedicine Online*, 25(6), pp.549–550. Elsevier.

Johnson, T.B. and Coghill, R.D., 1925. Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. *Journal of the American Chemical Society*, 47(11), pp.2838–2844.

Jones, P.L., Veenstra, G.J.C., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P., 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics*, 19(2), pp.187–191.

Jorgensen, B.G. and Ro, S., 2019. Role of DNA methylation in the development and differentiation of intestinal epithelial cells and smooth muscle cells. *Journal of Neurogastroenterology and Motility*, 25(3), pp.377–386. Korean Society of Neurogastroenterology and Motility.

Jung, G., Hernández-Illán, E., Moreira, L., Balaguer, F. and Goel, A., 2020. Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nature Reviews Gastroenterology & Hepatology*, 17(2), pp.111–130.

Kaaij, L.T.J., van de Wetering, M., Fang, F., Decato, B., Molaro, A., van de Werken, H.J.G., van Es, J.H., Schuijers, J., de Wit, E., de Laat, W., Hannon, G.J., Clevers, H.C., Smith, A.D. and Ketting, R.F., 2013. DNA methylation dynamics during intestinal stem cell differentiation reveals enhancers driving gene expression in the villus. *Genome Biology*, 14(5), p.R50.

Kafer, G.R., Li, X., Horii, T., Suetake, I., Tajima, S., Hatada, I., Mark, P., Correspondence, C. and Carlton, P.M., 2016. 5-Hydroxymethylcytosine Marks Sites of DNA Damage and Promotes Genome Stability. *CellReports*, 14, pp.1283–1292.

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M. and Saitou, M., 2013. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *The EMBO Journal*, 32(3), pp.340–353.

Kai, M., Niinuma, T., Kitajima, H., Yamamoto, E., Harada, T., Aoki, H., Maruyama,

R., Toyota, M., Sasaki, Y., Sugai, T., Tokino, T., Nakase, H. and Suzuki, H., 2016. TET1 Depletion Induces Aberrant CpG Methylation in Colorectal Cancer Cells. *PLOS ONE*, 11(12), p.e0168281.

Kane, A.E. and Sinclair, D.A., 2019. Epigenetic changes during aging and their reprogramming potential. *Critical Reviews in Biochemistry and Molecular Biology*, 54(1), pp.61–83.

Kang, J., Lienhard, M., Pastor, W.A., Chawla, A., Novotny, M., Tsagaratou, A., Lasken, R.S., Thompson, E.C., Azim Surani, M., Koralov, S.B., Kalantry, S., Chavez, L. and Rao, A., 2015. Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 112(31), pp.E4236–E4245.

Kang, J.H., Kim, S.J., Noh, D.Y., Park, I.A., Choe, K.J., Yoo, O.J. and Kang, H.S., 2001. Methylation in the p53 promoter is a supplementary route to breast carcinogenesis: Correlation between CpG methylation in the p53 promoter and the mutation of the p53 gene in the progression from ductal carcinoma in situ to invasive ductal carcinoma. *Laboratory Investigation*, 81(4), pp.573–579.

Karlen, Y., McNair, A., Perseguers, S., Mazza, C. and Mermod, N., 2007. Statistical significance of quantitative PCR. *BMC Bioinformatics*, 8, p.131.

Kazakevych, J., Sayols, S., Messner, B., Krienke, C. and Soshnikova, N., 2017. Dynamic changes in chromatin states during specification and differentiation of adult intestinal stem cells. *Nucleic Acids Research*, 45(10), pp.5770–5784.

Kellinger, M.W., Song, C.X., Chong, J., Lu, X.Y., He, C. and Wang, D., 2012. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nature Structural and Molecular Biology*, 19(8), pp.831–833.

Khaltourina, D., Matveyev, Y., Alekseev, A., Cortese, F. and Ioviţă, A., 2020. Aging Fits the Disease Criteria of the International Classification of Diseases. *Mechanisms of Ageing and Development*, 189, p.111230.

Khanna, C. and Hunter, K., 2009. Modeling metastasis in vivo. *Carcinogenesis*, 30(8),

pp.1330–1335.

Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M., Tice, R., Tochigi, M., Moréra, S., Nazarians, A., Belsham, D., Wong, A.H.C., Blencowe, B.J., Wang, S.C., Kapranov, P., Kustra, R., Labrie, V., Klimasauskas, S. and Petronis, A., 2012. 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nature Structural and Molecular Biology*, 19(10), pp.1037–1044.

Khavari, D.A., Sen, G.L. and Rinn, J.L., 2010. DNA methylation and epigenetic control of cellular differentiation. *Cell Cycle*, 9(19), pp.3880–3883. Taylor and Francis Inc.

Khoueiry, R., Sohni, A., Thienpont, B., Luo, X., Velde, J. Vande, Bartoccetti, M., Boeckx, B., Zwijsen, A., Rao, A., Lambrechts, D. and Koh, K.P., 2017. Lineage-specific functions of TET1 in the postimplantation mouse embryo. *Nature Genetics*, 49(7), pp.1061–1072.

Kim, I., Lee, J.W., Lee, M., Kim, H.S., Chung, H.H., Kim, J.W., Park, N.H., Song, Y.S. and Seo, J.S., 2018. Genomic landscape of ovarian clear cell carcinoma via whole exome sequencing. *Gynecologic Oncology*, 148(2), pp.375–382.

Kim, R., Sheaffer, K.L., Choi, I., Won, K.J. and Kaestner, K.H., 2016. Epigenetic regulation of intestinal stem cells by Tet1-mediated DNA hydroxymethylation. *Genes and Development*, 30(21), pp.2433–2442.

Kim, Y., Wang, S.E. and Jiang, Y. hui, 2019. Epigenetic therapy of Prader–Willi syndrome. *Translational Research*, 208, pp.105–118. Mosby Inc.

Kimura, Y., Shofuda, T., Higuchi, Y., Nagamori, I., Oda, M., Nakamori, M., Onodera, M., Kanematsu, D., Yamamoto, A., Katsuma, A., Suemizu, H., Nakano, T., Kanemura, Y. and Mochizuki, H., 2019. Human Genomic Safe Harbors and the Suicide Gene-Based Safeguard System for iPSC-Based Cell Therapy. *STEM CELLS Translational Medicine*, 8(7), p.sctm.18-0039.

Kitsera, N., Allgayer, J., Parsa, E., Geier, N., Rossa, M., Carell, T. and Khobta, A., 2017. Functional impacts of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine at a single hemi-modified CpG dinucleotide in a gene promoter.

Nucleic Acids Research, 45(19), pp.11033–11042.

Klemm, S.L., Shipony, Z. and Greenleaf, W.J., 2019. Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, 20(4), pp.207–220. Nature Publishing Group.

Klimant, E., Wright, H., Rubin, D., Seely, D. and Markman, M., 2018. Intravenous vitamin C in the supportive care of cancer patients: A review and rational approach. *Current Oncology*, 25(2), pp.139–148. Multimed Inc.

Ko, M., An, J., Bandukwala, H.S., Chavez, L., Äijö, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lähdesmäki, H., Hogan, P.G., Aravind, L. and Rao, A., 2013. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature*, 497(7447), pp.122–126.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., Liu, X.S., Aravind, L., Agarwal, S., MacIejewski, J.P. and Rao, A., 2010. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, 468(7325), pp.839–843.

Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., Sakashita, A., Wakai, T., Suzuki, Y., Ito, T., Matsui, Y. and Kono, T., 2013. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Research*, 23(4), pp.616–627.

Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G., Lahesmaa, R., Orkin, S.H., Rodig, S.J., Daley, G.Q. and Rao, A., 2011. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, 8(2), pp.200–213.

Kosmider, O., Gelsi-Boyer, V., Ciudad, M., Racoeur, C., Jooste, V., Vey, N., Quesnel, B., Fenaux, P., Bastie, J.N., Beyne-Rauzy, O., Stamatoulas, A., Dreyfus, F., Ifrah, N., De Botton, S., Vainchenker, W., Bernard, O.A., Birnbaum, D., Fontenay, M. and Solary, E., 2009. TET2 gene mutation is a frequent and adverse event in chronic myelomonocytic leukemia. *Haematologica*, 94(12), pp.1676–1681.

Kota, S.K. and Feil, R., 2010. Epigenetic Transitions in Germ Cell Development and

Meiosis. *Developmental Cell*, 19(5), pp.675–686. Elsevier.

Kotani, A., Kakazu, N., Tsuruyama, T., Okazaki, I.M., Muramatsu, M., Kinoshita, K., Nagaoka, H., Yabe, D. and Honjo, T., 2007. Activation-induced cytidine deaminase (AID) promotes B cell lymphomagenesis in Emu-cmyc transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America*, 104(5), pp.1616–1620.

Koutsouraki, E., Pells, S. and De Sousa, P.A., 2019. Sufficiency of hypoxia-inducible 2-oxoglutarate dioxygenases to block chemical oxidative stress-induced differentiation of human embryonic stem cells. *Stem Cell Research*, 34, p.101358.

Kremer, E.A., Gaur, N., Lee, M.A., Engmann, O., Bohacek, J. and Mansuy, I.M., 2018. Interplay between TETs and microRNAs in the adult brain for memory formation. *Scientific Reports*, 8(1), p.1678.

Kriaucionis, S. and Heintz, N., 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science*, 324(5929), pp.929–930.

Kubens, B.S. and Zänker, K.S., 1998. Differences in the migration capacity of primary human colon carcinoma cells (SW480) and their lymph node metastatic derivatives (SW620). *Cancer Letters*. Elsevier, pp.55–64.

Kudo, Y., Tateishi, K., Yamamoto, K., Yamamoto, S., Asaoka, Y., Ijichi, H., Nagae, G., Yoshida, H., Aburatani, H. and Koike, K., 2012. Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Science*, 103(4), pp.670–676.

Kumazaki, T., Takahashi, T., Matsuo, T., Kamada, M. and Mitsui, Y., 2013. Reemergence of undifferentiated cells from transplants of human induced pluripotent stem cells is a possible potential risk factor of tumorigenic differentiation. *Cell Biology International Reports*, 21(1), p.n/a-n/a.

Kundu, A., Shelar, S., Ghosh, A.P., Ballestas, M., Kirkman, R., Nam, H., Brinkley, G.J., Karki, S., Mobley, J.A., Bae, S., Varambally, S. and Sudarshan, S., 2020. 14-3-3 proteins protect AMPK-phosphorylated ten-eleven translocation-2 (TET2) from PP2A-mediated dephosphorylation. *Journal of Biological Chemistry*, 295(6), pp.1754–1766.

Kurrey, N.K., Jalgaonkar, S.P., Joglekar, A. V., Ghanate, A.D., Chaskar, P.D., Doiphode, R.Y. and Bapat, S.A., 2009. Snail and slug mediate radioresistance and chemoresistance by antagonizing p53-mediated apoptosis and acquiring a stem-like phenotype in ovarian cancer cells. *Stem Cells*, 27(9), pp.2059–2068.

Lam, K., Pan, K., Linnekamp, J.F., Medema, J.P. and Kandimalla, R., 2016. DNA methylation based biomarkers in colorectal cancer: A systematic review. *Biochimica et Biophysica Acta - Reviews on Cancer*, 1866(1), pp.106–120. Elsevier B.V.

Langlois, T., da Costa Reis Monte-Mor, B., Lenglet, G., Droin, N., Marty, C., Le Couédic, J.-P., Almire, C., Auger, N., Mercher, T., Delhommeau, F., Christensen, J., Helin, K., Debili, N., Fuks, F., Bernard, O.A., Solary, E., Vainchenker, W. and Plo, I., 2014. TET2 Deficiency Inhibits Mesoderm and Hematopoietic Differentiation in Human Embryonic Stem Cells. *STEM CELLS*, 32(8), pp.2084–2097.

Lasho, T.L., Vallapureddy, R., Finke, C.M., Mangaonkar, A., Gangat, N., Ketterling, R., Tefferi, A. and Patnaik, M.M., 2018. Infrequent occurrence of TET1, TET3, and ASXL2 mutations in myelodysplastic/myeloproliferative neoplasms. *Blood Cancer Journal*, 8(3), p.32. Nature Publishing Group.

Laukka, T., Mariani, C.J., Ihantola, T., Cao, J.Z., Hokkanen, J., Kaelin, W.G., Godley, L.A. and Koivunen, P., 2016. Fumarate and succinate regulate expression of hypoxia-inducible genes via TET enzymes. *Journal of Biological Chemistry*, 291(8), pp.4256–4265.

Lawson, K.A., Dunn, N.R., Roelen, B.A.J., Zeinstra, L.M., Davis, A.M., Wright, C.V.E., Korving, J.P.W.F.M. and Hogan, B.L.M., 1999. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes and Development*, 13(4), pp.424–436.

Le, T., Kim, K.P., Fan, G. and Faull, K.F., 2011. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. *Analytical Biochemistry*, 412(2), pp.203–209.

Lee, H.J., Hore, T.A. and Reik, W., 2014. Reprogramming the methylome: Erasing memory and creating diversity. *Cell Stem Cell*, 14(6), pp.710–719. Cell Press.

Lee, J., Bignone, P.A., Coles, L.S., Liu, Y., Snyder, E. and Larocca, D., 2020. Induced pluripotency and spontaneous reversal of cellular aging in supercentenarian donor cells: Induced Pluripotent Stem Cells from a 114-year-old Supercentenarian. *Biochemical and Biophysical Research Communications*, 525(3), pp.563–569.

Lei, Y., Huang, K., Gao, C., Lau, Q.C., Pan, H., Xie, K., Li, J., Liu, R., Zhang, T., Xie, N., Nai, H.S., Wu, H., Dong, Q., Zhao, X., Nice, E.C., Huang, C. and Wei, Y., 2011. Proteomics identification of ITGB3 as a key regulator in reactive oxygen species-induced migration and invasion of colorectal cancer cells. *Molecular and Cellular Proteomics*, 10(10).

Leibovitz, A., Stinson, J.C., McCombs, W.B., McCoy, C.E., Mazur, K.C. and Mabry, N.D., 1976. Human Colorectal Adenocarcinoma Cell Lines. *Cancer Research*, 36, pp.4562–4569.

Leitch, H.G., Mcewen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., Surani, M.A. and Hajkova, P., 2013. Naive pluripotency is associated with global DNA hypomethylation. *Nature Structural and Molecular Biology*, 20(3), pp.311–316.

Leitch, H.G., Tang, W.W.C. and Surani, M.A., 2013. Primordial Germ-Cell Development and Epigenetic Reprogramming in Mammals. *Current Topics in Developmental Biology*. Academic Press Inc., pp.149–187.

Lemonnier, F., Couronné, L., Parrens, M., Jaïs, J.P., Travert, M., Lamant, L., Tournillac, O., Rousset, T., Fabiani, B., Cairns, R.A., Mak, T., Bastard, C., Bernard, O.A., De Leval, L. and Gaulard, P., 2012. Recurrent TET2 mutations in peripheral T-cell lymphomas correlate with T FH-like features and adverse clinical parameters. *Blood*, 120(7), pp.1466–1469.

Lemonnier, F., Poullot, E., Dupuy, A., Couronné, L., Martin, N., Scourzic, L., Fataccioli, V., Bruneau, J., Cairns, R.A., Mak, T.W., Bernard, O.A., de Leval, L. and Gaulard, P., 2018. Loss of 5-hydroxymethylcytosine is a frequent event in peripheral T-cell lymphomas. *Haematologica*, 103(3), pp.e115–e118. Ferrata Storti Foundation.

Leoz, M.L., Carballal, S., Moreira, L., Ocaña, T. and Balaguer, F., 2015. The genetic basis of familial adenomatous polyposis and its implications for clinical practice

and risk management. *Application of Clinical Genetics*, 8, pp.95–107. Dove Medical Press Ltd.

Letouzé, E., Martinelli, C., Loriot, C., Burnichon, N., Abermil, N., Ottolenghi, C., Janin, M., Menara, M., Nguyen, A.T., Benit, P., Buffet, A., Marcaillou, C., Bertherat, J.Ô., Amar, L., Rustin, P., DeReyniès, A., Gimenez-Roqueplo, A.P. and Favier, J., 2013. SDH Mutations Establish a Hypermethylator Phenotype in Paraganglioma. *Cancer Cell*, 23(6), pp.739–752.

Levine, M.E., Hosgood, H.D., Chen, B., Absher, D., Assimes, T. and Horvath, S., 2015. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging*, 7(9), pp.690–700.

Levine, M.E., Lu, A.T., Bennett, D.A. and Horvath, S., 2015. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging*, 7(12), pp.1198–1211.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A., 1992. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to Methylated DNA. *Cell*, 69(6), pp.905–914.

Ley, T.J., Miller, C., Ding, L., Raphael, B.J., Mungall, A.J., Robertson, G., Hoadley, K., Triche, T.J., Laird, P.W., Baty, J.D., Fulton, L.L., Fulton, R., Heath, S.E., Kalicki-Veizer, J., Kandoth, C., Klco, J.M., Koboldt, D.C., Kanchi, K.L., Kulkarni, S., Lamprecht, T.L., Larson, D.E., Lin, G., Lu, C., McLellan, M.D., McMichael, J.F., Payton, J., Schmidt, H., Spencer, D.H., Tomasson, M.H., Wallis, J.W., Wartman, L.D., Watson, M.A., Welch, J., Wendl, M.C., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y., Chiu, R., Chu, A., Chuah, E., Chun, H.J., Corbett, R., Dhalla, N., Guin, R., He, A., Hirst, C., Hirst, M., Holt, R.A., Jones, S., Karsan, A., Lee, D., Li, H.I., Marra, M.A., Mayo, M., Moore, R.A., Mungall, K., Parker, J., Pleasance, E., Plettner, P., Schein, J., Stoll, D., Swanson, L., Tam, A., Thiessen, N., Varhol, R., Wye, N., Zhao, Y., Gabriel, S., Getz, G., Sougnez, C., Zou, L., Leiserson, M.D.M., Vandin, F., Wu, H.T., Applebaum, F., Baylin, S.B., Akbani, R., Broom, B.M., Chen, K., Motter, T.C., Nguyen, K., Weinstein, J.N., Zhang, N., Ferguson, M.L., Adams, C., Black, A., Bowen, J., Gastier-Foster, J., Grossman, T., Lichtenberg, T., Wise, L., Davidsen, T., Demchok, J.A., Mills Shaw, K.R., Sheth, M., Sofia, H.J., Yang, L., Downing, J.R., Eley, G., Alonso, S., Ayala, B., Baboud,

J., Backus, M., Barletta, S.P., Berton, D.L., Chu, A.L., Girshik, S., Jensen, M.A., Kahn, A., Kothiyal, P., Nicholls, M.C., Pihl, T.D., Pot, D.A., Raman, R., Sanbhadti, R.N., Snyder, E.E., Srinivasan, D., Walton, J., Wan, Y., Wang, Z., Issa, J.P.J., Beau, M. Le, Carroll, M., Kantarjian, H., Kornblau, S., Bootwalla, M.S., Lai, P.H., Shen, H., Van Den Berg, D.J., Weisenberger, D.J., Link, D.C., Walter, M.J., Ozenberger, B.A., Mardis, E.R., Westervelt, P., Graubert, T.A., DiPersio, J.F. and Wilson, R.K., 2013. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *New England Journal of Medicine*, 368(22), pp.2059–2074.

Li, E., Bestor, T.H. and Jaenisch, R., 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6), pp.915–926.

Li, F., Zhang, Y., Bai, J., Greenberg, M.M., Xi, Z. and Zhou, C., 2017. 5-Formylcytosine Yields DNA-Protein Cross-Links in Nucleosome Core Particles. *Journal of the American Chemical Society*, 139(31), pp.10617–10620.

Li, H., Zhou, Z.Q., Yang, Z.R., Tong, D.N., Guan, J., Shi, B.J., Nie, J., Ding, X.T., Li, B., Zhou, G.W. and Zhang, Z.Y., 2017. MicroRNA-191 acts as a tumor promoter by modulating the TET1–p53 pathway in intrahepatic cholangiocarcinoma. *Hepatology*, 66(1), pp.136–151.

Li, J, Wu, X., Zhou, Y., Lee, M., Guo, L., … W.H.-N. acids and 2018, undefined, 2018. Decoding the dynamic DNA methylation and hydroxymethylation landscapes in endodermal lineage intermediates during pancreatic differentiation of hESC. *Nucleic Acids Research*, 46(6), pp.2883–2900.

Li, Jia, Wu, X., Zhou, Y., Lee, M., Guo, L., Han, W., Mo, W., Cao, W.-M., Sun, D., Xie, R. and Huang, Y., 2018. Decoding the dynamic DNA methylation and hydroxymethylation landscapes in endodermal lineage intermediates during pancreatic differentiation of hESC. *Nucleic Acids Research*, 46(6), pp.2883–2900.

Li, L., Li, C., Mao, H., Du, Z., Chan, W.Y., Murray, P., Luo, B., Chan, A.T.C., Mok, T.S.K., Chan, F.K.L., Ambinder, R.F. and Tao, Q., 2016. Epigenetic inactivation of the CpG demethylase TET1 as a DNA methylation feedback loop in human cancers. *Scientific Reports*, 6(1), pp.1–13.

Li, L., Li, R. and Wang, Y., 2019. Identification of selective and reversible LSD1

inhibitors with anti-metastasis activity by high-throughput docking. *Bioorganic and Medicinal Chemistry Letters*, 29(4), pp.544–548.

Li, T., Wang, L., Du, Y., Xie, S., Yang, X., Lian, F., Zhou, Z. and Qian, C., 2018. Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. *Nucleic Acids Research*, 46(6), pp.3218–3231.

Li, W. and Liu, M., 2011. Distribution of 5-Hydroxymethylcytosine in Different Human Tissues. *Research Journal of Nucleic Acids*, 2011.

Li, W., Zhang, Xu, Lu, X., You, L., Song, Y., Luo, Z., Zhang, J., Nie, J., Zheng, W., Xu, D., Wang, Y., Dong, Y., Yu, S., Hong, J., Shi, J., Hao, H., Luo, F., Hua, L., Wang, P., Qian, X., Yuan, F., Wei, L., Cui, M., Zhang, T., Liao, Q., Dai, M., Liu, Z., Chen, G., Meckel, K., Adhikari, S., Jia, G., Bissonnette, M.B., Zhang, Xinxiang, Zhao, Y., Zhang, W., He, C. and Liu, J., 2017. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Research*, 27(10), pp.1243–1257.

Li, Xin, Liu, Y., Salz, T., Hansen, K.D. and Feinberg, A., 2016. Whole-genome analysis of the methylome and hydroxymethylome in normal and malignant lung and liver. *Genome Research*, 26(12), pp.1730–1741.

Li, X., Xiao, B. and Chen, X.S., 2017. DNA Methylation: a New Player in Multiple Sclerosis. *Molecular Neurobiology*, 54(6), pp.4049–4059. Humana Press Inc.

Li, Xiang, Yue, X., Pastor, W.A., Lin, L., Georges, R., Chavez, L., Evans, S.M. and Rao, A., 2016. Tet proteins influence the balance between neuroectodermal and mesodermal fate choice by inhibiting Wnt signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 113(51), pp.E8267–E8276.

Li, Z., Cai, X., Cai, C.L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.C. and Xu, M., 2011. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood*. Blood, pp.4509–4518.

Lian, C.G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., Lee, C.W., Hu, D., Lian, B.Q., Kleffel, S., Yang, Y., Neiswender, J., Khorasani, A.J., Fang, R., Lezcano, C., Duncan, L.M., Scolyer, R.A., Thompson, J.F., Kakavand, H., Houvras, Y., Zon, L.I., Mihm, M.C., Kaiser, U.B., Schatton,

T., Woda, B.A., Murphy, G.F. and Shi, Y.G., 2012. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of Melanoma. *Cell*, 150(6), pp.1135–1146.

Liao, Y., Gu, J., Wu, Y., Long, X., Ge, D., Xu, J. and Ding, J., 2016. Level of 5-hydroxymethylcytosine predicts poor prognosis in non-small cell lung cancer. *Oncology Letters*, 11(6), pp.3753–3760.

Lichtenstern, C.R., Ngu, R.K., Shalapour, S. and Karin, M., 2020. Immunotherapy, Inflammation and Colorectal Cancer. *Cells*, 9(3), p.618.

Lin, L.L., Wang, W., Hu, Z.Y., Wang, L.W., Chang, J. and Qian, H.G., 2015. Erratum to: Negative feedback of miR-29 family TET1 involves in hepatocellular cancer. *Medical Oncology*, 32(3), pp.1–11. Humana Press Inc.

Lin, M.H., Chang, K.W., Lin, S.C. and Miner, J.H., 2010. Epidermal hyperproliferation in mice lacking fatty acid transport protein 4 (FATP4) involves ectopic EGF receptor and STAT3 signaling. *Developmental Biology*, 344(2), pp.707–719.

Lio, C.-W.J., Yuita, H. and Rao, A., 2019. Dysregulation of the TET family of epigenetic regulators in lymphoid and myeloid malignancies. *Blood*, 134(18), pp.1487–1497.

Lio, C.W.J. and Rao, A., 2019. TET enzymes and 5hMC in adaptive and innate immune systems. *Frontiers in Immunology*, 10(FEB). Frontiers Media S.A.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J.C., Rao, A., Esteller, M., He, C., Haghighi, F.G., Sejnowski, T.J., Behrens, M.M. and Ecker, J.R., 2013. Global epigenomic reconfiguration during mammalian brain development. *Science*, 341(6146).

Liu, D., Li, G. and Zuo, Y., 2019. Function determinants of TET proteins: The arrangements of sequence motifs with specific codes. *Briefings in Bioinformatics*, 20(5), pp.1826–1835. Oxford University Press.

Liu, H., Xu, T., Cheng, Y., Jin, M.H., Chang, M.Y., Shu, Q., Allen, E.G., Jin, P. and Wang, X., 2019. Altered 5-Hydroxymethylcytosine Landscape in Primary Gastric Adenocarcinoma. *DNA and Cell Biology*, 38(12), pp.1460–1469.

Liu, L., Wang, H.J., Meng, T., Lei, C., Yang, X.H., Wang, Q.S., Jin, B. and Zhu, J.F., 2019. lncRNA GAS5 Inhibits Cell Migration and Invasion and Promotes Autophagy by Targeting miR-222-3p via the GAS5/PTEN-Signaling Pathway in CRC. *Molecular Therapy - Nucleic Acids*, 17, pp.644–656.

Liu, N., Wang, M., Deng, W., Schmidt, C.S., Qin, W., Leonhardt, H. and Spada, F., 2013. Intrinsic and Extrinsic Connections of Tet3 Dioxygenase with CXXC Zinc Finger Modules. *PLoS ONE*, 8(5).

Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H. and Wong, J., 2013. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nature Communications*, 4.

Liu, Y., Aryee, M.J., Padyukov, L., Daniele Fallin, M., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., Shchetynsky, K., Scheynius, A., Kere, J., Alfredsson, L., Klareskog, L., Ekstr, T.J. and Feinberg, A.P., 2013. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis.

Liu, Y., Toh, H., Sasaki, H., Zhang, X. and Cheng, X., 2012. An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. *Genes and Development*, 26(21), pp.2374–2379.

Liu, Z., Feng, Q., Sun, P., Lu, Y., Yang, M., Zhang, X., Jin, X., Li, Y., Lu, S.-J. and Quan, C., 2017. Genome-wide DNA methylation drives human embryonic stem cell erythropoiesis by remodeling gene expression dynamics. *Epigenomics*, 9(12), pp.1543–1558.

Livak, K.J. and Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*, 25(4), pp.402–408.

Lorsback, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T. and Downing, J.R., 2003. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;23) [3]. *Leukemia*, 17(3), pp.637–641. Nature Publishing Group.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. and Zhang, Y., 2014. Role of Tet proteins in enhancer activity and telomere elongation. *Genes & Development*, 28(19),

pp.2103–2119.

Lu, X., Han, D., Zhao, B.S., Song, C.X., Zhang, L.S., Doré, L.C. and He, C., 2015. Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. *Cell Research*, 25(3), pp.386–389. Nature Publishing Group.

Luo, J., Li, Y.N., Wang, F., Zhang, W.M. and Geng, X., 2010. S-adenosylmethionine inhibits the growth of cancer cells by reversing the hypomethylation status of c-myc and H-ras in human gastric cancer and colon cancer. *International Journal of Biological Sciences*, 6(7), pp.784–795.

Lustgarten, M.S., 2016. Classifying Aging As a Disease: The Role of Microbes. *Frontiers in Genetics*, 7(DEC), p.212.

Lv, X., Jiang, H., Liu, Y., Lei, X. and Jiao, J., 2014. Micro <scp>RNA</scp> -15b promotes neurogenesis and inhibits neural progenitor proliferation by directly repressing <scp>TET</scp> 3 during early neocortical development. *EMBO reports*, 15(12), pp.1305–1314.

Lykkesfeldt, J. and Tveden-Nyborg, P., 2019. The pharmacokinetics of vitamin C. *Nutrients*, 11(10). MDPI AG.

Ma, C., Rong, Y., Radiloff, D.R., Datto, M.B., Centeno, B., Bao, S., Cheng, A.W.M., Lin, F., Jiang, S., Yeatman, T.J. and Wang, X.F., 2008. Extracellular matrix protein βig-h3/TGFBI promotes metastasis of colon cancer by enhancing cell extravasation. *Genes and Development*, 22(3), pp.308–321.

Ma, S., Chan, K.W., Hu, L., Lee, T.K.W., Wo, J.Y.H., Ng, I.O.L., Zheng, B.J. and Guan, X.Y., 2007. Identification and Characterization of Tumorigenic Liver Cancer Stem/Progenitor Cells. *Gastroenterology*, 132(7), pp.2542–2556.

Ma, Y., Qin, C., Li, L., Miao, R., Jing, C. and Cui, X., 2018. MicroRNA-21 promotes cell proliferation by targeting tumor suppressor TET1 in colorectal cancer. *International journal of clinical and experimental pathology*, 11(3), pp.1439–1445.

Macaluso, M., Paggi, M.G. and Giordano, A., 2003. Genetic and epigenetic alterations as hallmarks of the intricate road to cancer. *Oncogene*, 22(43), pp.6472–6478. Nature Publishing Group.

Mahaddalkar, P.U., Scheibner, K., Pfluger, S., Ansarullah, Sterr, M., Beckenbauer, J., Irmler, M., Beckers, J., Knöbel, S. and Lickert, H., 2020. Generation of pancreatic β cells from CD177+ anterior definitive endoderm. *Nature Biotechnology*, pp.1–12.

Mahé, E.A., Madigou, T., Sérandour, A.A., Bizot, M., Avner, S., Chalmel, F., Palierne, G., Métivier, R. and Salbert, G., 2017. Cytosine modifications modulate the chromatin architecture of transcriptional enhancers. *Genome Research*, 27(6), pp.947–958.

Maiti, A. and Drohat, A.C., 2011. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of CpG sites. *Journal of Biological Chemistry*, 286(41), pp.35334–35338.

Maltby, V.E., Lea, R.A., Graves, M.C., Sanders, K.A., Benton, M.C., Tajouri, L., Scott, R.J. and Lechner-Scott, J., 2018. Genome-wide DNA methylation changes in CD19+ B cells from relapsing-remitting multiple sclerosis patients. *Scientific Reports*, 8(1), p.17418.

Mannavola, F., D'oronzo, S., Cives, M., Stucci, L.S., Ranieri, G., Silvestris, F. and Tucci, M., 2020. Extracellular vesicles and epigenetic modifications are hallmarks of melanoma progression. *International Journal of Molecular Sciences*, 21(1), p.52. MDPI AG.

Marçais, A., Waast, L., Bruneau, J., Hanssens, K., Asnafi, V., Gaulard, P., Suarez, F., Dubreuil, P., Gessain, A., Hermine, O. and Pique, C., 2017. Adult T cell leukemia aggressivenness correlates with loss of both 5-hydroxymethylcytosine and TET2 expression. *Oncotarget*, 8(32), pp.52256–52268.

Margalit, S., Avraham, S., Shahal, T., Michaeli, Y., Gilat, N., Magod, P., Caspi, M., Loewenstein, S., Lahat, G., Friedmann-Morvinski, D., Kariv, R., Rosin-Arbesfeld, R., Zirkin, S. and Ebenstein, Y., 2020. 5-Hydroxymethylcytosine as a clinical biomarker: Fluorescence-based assay for high-throughput epigenetic quantification in human tissues. *International Journal of Cancer*, 146(1), pp.115–122.

Mariadason, J.M., 2008. HDACs and HDAC inhibitors in colon cancer. *Epigenetics*, 3(1), pp.28–37.

Mariadason, J.M., Nicholas, C., L'Italien, K.E., Zhuang, M., Smartt, H.J.M., Heerdt, B.G., Yang, W., Corner, G.A., Wilson, A.J., Klampfer, L., Arango, D. and Augenlicht, L.H., 2005. Gene expression profiling of intestinal epithelial cell maturation along the crypt-villus axis. *Gastroenterology*, 128(4), pp.1081–1088.

Marina, R.J., Sturgill, D., Bailly, M.A., Thenoz, M., Varma, G., Prigge, M.F., Nanan, K.K., Shukla, S., Haque, N. and Oberdoerffer, S., 2016. TET -catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF -dependent alternative splicing . *The EMBO Journal*, 35(3), pp.335–355.

Marioni, R.E., Shah, S., McRae, A.F., Chen, B.H., Colicino, E., Harris, S.E., Gibson, J., Henders, A.K., Redmond, P., Cox, S.R., Pattie, A., Corley, J., Murphy, L., Martin, N.G., Montgomery, G.W., Feinberg, A.P., Fallin, M.D., Multhaup, M.L., Jaffe, A.E., Joehanes, R., Schwartz, J., Just, A.C., Lunetta, K.L., Murabito, J.M., Starr, J.M., Horvath, S., Baccarelli, A.A., Levy, D., Visscher, P.M., Wray, N.R. and Deary, I.J., 2015. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biology*, 16(1), p.25.

Mármol, I., Sánchez-de-Diego, C., Dieste, A.P., Cerrada, E. and Yoldi, M.J.R., 2017. Colorectal carcinoma: A general overview and future perspectives in colorectal cancer. *International Journal of Molecular Sciences*, 18(1). MDPI AG.

Martin, R.M., Fowler, J.L., Cromer, M.K., Lesch, B.J., Ponce, E., Uchida, N., Nishimura, T., Porteus, M.H. and Loh, K.M., 2020. Improving the safety of human pluripotent stem cell therapies using genome-edited orthogonal safeguards. *Nature Communications*, 11(1), pp.1–14.

McBrayer, S.K., Mayers, J.R., DiNatale, G.J., Shi, D.D., Khanal, J., Chakraborty, A.A., Sarosiek, K.A., Briggs, K.J., Robbins, A.K., Sewastianik, T., Shareef, S.J., Olenchock, B.A., Parker, S.J., Tateishi, K., Spinelli, J.B., Islam, M., Haigis, M.C., Looper, R.E., Ligon, K.L., Bernstein, B.E., Carrasco, R.D., Cahill, D.P., Asara, J.M., Metallo, C.M., Yennawar, N.H., Vander Heiden, M.G. and Kaelin, W.G., 2018. Transaminase Inhibition by 2-Hydroxyglutarate Impairs Glutamate Biosynthesis and Redox Homeostasis in Glioma. *Cell*, 175(1), pp.101-116.e25.

Medvedeva, Y.A., Khamis, A.M., Kulakovskiy, I. V., Ba-Alawi, W., Bhuyan, M.S.I., Kawaji, H., Lassmann, T., Harbers, M., Forrest, A.R.R. and Bajic, V.B., 2014.

Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics*, 15(1), pp.1–12.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L. and Bird, A.P., 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*, 58(3), pp.499–507.

Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. and Heintz, N., 2012. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*, 151(7), pp.1417–1430.

Menezes, J., Acquadro, F., Wiseman, M., Gómez-López, G., Salgado, R.N., Talavera-Casañas, J.G., Buño, I., Cervera, J. V., Montes-Moreno, S., Hernández-Rivas, J.M., Ayala, R., Calasanz, M.J., Larrayoz, M.J., Brichs, L.F., Gonzalez-Vicent, M., Pisano, D.G., Piris, M.A., Álvarez, S. and Cigudosa, J.C., 2014. Exome sequencing reveals novel and recurrent mutations with clinical impact in blastic plasmacytoid dendritic cell neoplasm. *Leukemia*, 28(4), pp.823–829.

Meran, L., Baulies, A. and Li, V.S.W., 2017. Intestinal Stem Cell Niche: The Extracellular Matrix and Cellular Components. *Stem Cells International*, 2017. Hindawi Limited.

Messerschmidt, D.M., De Vries, W., Ito, M., Solter, D., Ferguson-Smith, A. and Knowles, B.B., 2012. Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science*, 335(6075), pp.1499–1502.

Minor, E.A., Court, B.L., Young, J.I. and Wang, G., 2013. Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *Journal of Biological Chemistry*, 288(19), pp.13669–13674.

Mirabelli, P., Coppola, L. and Salvatore, M., 2019. Cancer cell lines are useful model systems for medical research. *Cancers*, 11(8). MDPI AG.

Misawa, K., Yamada, S., Mima, M., Nakagawa, T., Kurokawa, T., Imai, A., Mochizuki, D., Morita, K., Ishikawa, R., Endo, S. and Misawa, Y., 2019. 5-Hydroxymethylcytosine and ten-eleven translocation dioxygenases in head and neck carcinoma. *Journal of Cancer*, 10(21), pp.5306–5314.

Misiakos, E.P., Karidis, N.P. and Kouraklis, G., 2011. Current treatment for colorectal liver metastases. *World Journal of Gastroenterology*, 17(36), pp.4067–4075.

Mo, H.Y., An, C.H., Choi, E.J., Yoo, N.J. and Lee, S.H., 2020. Somatic mutation and loss of expression of a candidate tumor suppressor gene TET3 in gastric and colorectal cancers. *Pathology Research and Practice*, 216(3), p.152759.

Mongelli, A., Martelli, F., Farsetti, A. and Gaetano, C., 2019. The dark that matters: Long noncoding RNAs as master regulators of cellular metabolism in noncommunicable diseases. *Frontiers in Physiology*, 10(MAY), p.369. Frontiers Media S.A.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W. and Petersen-Mahrt, S.K., 2004. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: Implications for epigenetic reprogramming. *Journal of Biological Chemistry*, 279(50), pp.52353–52360.

Morris, J.P., Yashinskie, J.J., Koche, R., Chandwani, R., Tian, S., Chen, C.C., Baslan, T., Marinkovic, Z.S., Sánchez-Rivera, F.J., Leach, S.D., Carmona-Fontaine, C., Thompson, C.B., Finley, L.W.S. and Lowe, S.W., 2019. α-Ketoglutarate links p53 to cell fate during tumour suppression. *Nature*, 573(7775), pp.595–599.

Morrison, J.R., Pászty, C., Stevens, M.E., Hughes, S.D., Forte, T., Scott, J. and Rubin, E.M., 1996. Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 93(14), pp.7154–7159.

Mulholland, C., Traube, F., Parsa, E., Eckl, E.-M., Schönung, M., Modic, M., Bartoschek, M., Stolz, P., Ryan, J., Carell, T., Leonhardt, H. and Bultmann, S., 2018. Distinct and stage-specific contributions of TET1 and TET2 to stepwise cytosine oxidation in the transition from naive to primed pluripotency. *bioRxiv*, p.281519.

Mummert, S.K., Lobanenkov, V.A. and Feinberg, A.P., 2005. Association of chromosome arm 16q loss with loss of imprinting of insulin-like growth factor-II in Wilms tumor. *Genes, Chromosomes and Cancer*, 43(2), pp.155–161.

Munari, E., Chaux, A., Vaghasia, A.M., Taheri, D., Karram, S., Bezerra, S.M.,

Gonzalez Roibon, N., Nelson, W.G., Yegnasubramanian, S., Netto, G.J. and Haffner, M.C., 2016. Global 5-Hydroxymethylcytosine Levels Are Profoundly Reduced in Multiple Genitourinary Malignancies. *PLOS ONE*, 11(1), p.e0146302.

Múnera, J.O., Sundaram, N., Rankin, S.A., Hill, D., Watson, C., Mahe, M., Vallance, J.E., Shroyer, N.F., Sinagoga, K.L., Zarzoso-Lacoste, A., Hudson, J.R., Howell, J.C., Chatuvedi, P., Spence, J.R., Shannon, J.M., Zorn, A.M., Helmrath, M.A. and Wells, J.M., 2017. Differentiation of Human Pluripotent Stem Cells into Colonic Organoids via Transient Activation of BMP Signaling. *Cell Stem Cell*, 21(1), pp.51-64.e6.

Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmiller, V., Michalakis, S., Müller, M., Biel, M. and Carell, T., 2010. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angewandte Chemie - International Edition*, 49(31), pp.5375–5377.

Münzel, M., Globisch, D. and Carell, T., 2011. 5-hydroxymethylcytosine, the sixth base of the genome. *Angewandte Chemie - International Edition*, 50(29), pp.6460–6468. Angew Chem Int Ed Engl.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y. and Honjo, T., 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, 102(5), pp.553–563.

Murata, A., Baba, Y., Ishimoto, T., Miyake, K., Kosumi, K., Harada, K., Kurashige, J., Iwagami, S., Sakamoto, Y., Miyamoto, Y., Yoshida, N., Yamamoto, M., Oda, S., Watanabe, M., Nakao, M. and Baba, H., 2015. TET family proteins and 5-hydroxymethylcytosine in esophageal squamous cell carcinoma. *Oncotarget*, 6(27), pp.23372–23382.

Murugan, A.K., Bojdani, E. and Xing, M., 2010. Identification and functional characterization of isocitrate dehydrogenase 1 (IDH1) mutations in thyroid cancer. *Biochemical and Biophysical Research Communications*, 393(3), pp.555–559.

Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y. and Kohli, R.M., 2012. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nature Chemical Biology*, 8(9), pp.751–758.

Nakagawa, T., Lv, L., Nakagawa, M., Yu, Y., Yu, C., D'Alessio, A.C., Nakayama, K., Fan, H.Y., Chen, X. and Xiong, Y., 2015. CRL4VprBP E3 ligase promotes monoubiquitylation and chromatin binding of TET dioxygenases. *Molecular Cell*, 57(2), pp.247–260.

Nakamura, N. and Takenaga, K., 1998. Hypomethylation of the metastasis-associated S100A4 gene correlates with gene activation in human colon adenocarcinoma cell lines. *Clinical and Experimental Metastasis*, 16(5), pp.471–479.

Nakamura, T., Liu, Y.J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y. and Nakano, T., 2012. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature*, 486(7403), pp.415–419.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683), pp.386–389.

Neri, F., Dettori, D., Incarnato, D., Krepelova, A., Rapelli, S., Maldotti, M., Parlato, C., Paliogiannis, P. and Oliviero, S., 2015. TET1 is a tumour suppressor that inhibits colon cancer growth by derepressing inhibitors of the WNT pathway. *Oncogene*, 34(32), pp.4168–4176.

Neri, Francesco, Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., Medana, C., DalBello, F. and Oliviero, S., 2015. Single-Base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA Methylation Dynamics. *Cell Reports*, 10(5), pp.674–683.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C. and Oliviero, S., 2013. Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biology*, 14(8), p.R91.

Nestor, C.E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., Katz, E., Dixon, J.M., Harrison, D.J. and Meehan, R.R., 2012. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Research*, 22(3), pp.467–477.

Ni, W., Yao, S., Zhou, Y., Liu, Y., Huang, P., Zhou, A., Liu, J., Che, L. and Li, J., 2019. Long noncoding RNA GAS5 inhibits progression of colorectal cancer by interacting with and triggering YAP phosphorylation and degradation and is negatively regulated by the m6A reader YTHDF3. *Molecular Cancer*, 18(1), p.143.

Nicholls, R.D. and Knepper, J.L., 2001. Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annual Review of Genomics and Human Genetics*, 2, pp.153–175. Annu Rev Genomics Hum Genet.

Nickerson, M.L., Im, K.M., Misner, K.J., Tan, W., Lou, H., Gold, B., Wells, D.W., Bravo, H.C., Fredrikson, K.M., Harkins, T.T., Milos, P., Zbar, B., Linehan, W.M., Yeager, M., Andresson, T., Dean, M. and Bova, G.S., 2013. Somatic Alterations Contributing to Metastasis of a Castration-Resistant Prostate Cancer. *Human Mutation*, 34(9), pp.1231–1241.

Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boué, J., Bertheas, M.F. and Mandel, J.L., 1991. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*, 252(5009), pp.1097–1102.

Oda, M., Kumaki, Y., Shigeta, M., Jakt, L.M., Matsuoka, C., Yamagiwa, A., Niwa, H. and Okano, M., 2013. DNA Methylation Restricts Lineage-specific Functions of Transcription Factor Gata4 during Embryonic Stem Cell Differentiation. *PLoS Genetics*, 9(6).

Odejide, O., Weigert, O., Lane, A.A., Toscano, D., Lunning, M.A., Kopp, N., Kim, S., Van Bodegom, D., Bolla, S., Schatz, J.H., Teruya-Feldstein, J., Hochberg, E., Louissaint, A., Dorfman, D., Stevenson, K., Rodig, S.J., Piccaluga, P.P., Jacobsen, E., Pileri, S.A., Harris, N.L., Ferrero, S., Inghirami, G., Horwitz, S.M. and Weinstock, D.M., 2014. A targeted mutational landscape of angioimmunoblastic T-cell lymphoma. *Blood*, 123(9), pp.1293–1296.

Ogaki, S., Shiraki, N., Kume, K. and Kume, S., 2013. Wnt and Notch signals guide embryonic stem cell differentiation into the intestinal lineages. *Stem Cells*, 31(6), pp.1086–1096.

Ogino, S., Kawasaki, T., Nosho, K., Ohnishi, M., Suemoto, Y., Kirkner, G.J. and Fuchs,

C.S., 2008. LINE-1 hypomethylation is inversely associated with microsatellite instability and CpG island methylator phenotype in colorectal cancer. *International Journal of Cancer*, 122(12), pp.2767–2773.

Ohno, S., Ohno, Y., Suzuki, N., Soma, G.I. and Inoue, M., 2009. High-dose vitamin C (ascorbic acid) therapy in the treatment of patients with advanced cancer. *Anticancer Research*. International Institute of Anticancer Research, pp.809–816.

Ohtani-Fujita, N., Fujita, T., Aoike, A., Osifchin, N.E., Robbins, P.D. and Sakai, T., 1993. CpG methylation inactivates the promoter activity of the human retinoblastoma tumor-suppressor gene. *Oncogene*, 8(4), pp.1063–1067.

Okamoto, Y., Yoshida, N., Suzuki, T., Shimozawa, N., Asami, M., Matsuda, T., Kojima, N., Perry, A.C.F. and Takada, T., 2016. DNA methylation dynamics in mouse preimplantation embryos revealed by mass spectrometry. *Scientific Reports*, 6(1), pp.1–9.

Okano, M., Bell, D.W., Haber, D.A. and Li, E., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3), pp.247–257.

Okano, M., Xie, S. and Li, E., 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases [1]. *Nature Genetics*, 19(3), pp.219–220. Nature Publishing Group.

Okano, Masaki, Xie, S. and Li, E., 1998. Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Research*, 26(11), pp.2536–2540.

Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H. and Hayashi, Y., 2002. LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Research*, 62(14), pp.4075–4080.

Oshimo, Y., Nakayama, H., Ito, R., Kitadai, Y., Yoshida, K., Chayama, K. and Yasui, W., 2003. Promoter methylation of cyclin D2 gene in gastric carcinoma. *International journal of oncology*, 23(6), pp.1663–1670.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik,

W. and Walter, J., 2000. Active demethylation of the paternal genome in the mouse zygote. *Current Biology*, 10(8), pp.475–478.

Otani, J., Kimura, H., Sharif, J., Endo, T.A., Mishima, Y., Kawakami, T., Koseki, H., Shirakawa, M., Suetake, I. and Tajima, S., 2013. Cell Cycle-Dependent Turnover of 5-Hydroxymethyl Cytosine in Mouse Embryonic Stem Cells. *PLoS ONE*, 8(12), p.e82961.

Ouakrim, D.A., Pizot, C., Boniol, Magali, Malvezzi, M., Boniol, Mathieu, Negri, E., Bota, M., Jenkins, M.A., Bleiberg, H. and Autier, P., 2015. Trends in colorectal cancer mortality in Europe: Retrospective analysis of the WHO mortality database. *The BMJ*, 351.

Ozawa, T., Matsuyama, T., Toiyama, Y., Takahashi, N., Ishikawa, T., Uetake, H., Yamada, Y., Kusunoki, M., Calin, G. and Goel, A., 2017. CCAT1 and CCAT2 long noncoding RNAs, located within the 8q.24.21 'gene desert', serve as important prognostic biomarkers in colorectal cancer. *Annals of Oncology*, 28(8), pp.1882–1888.

Palomero, T., Couronné, L., Khiabanian, H., Kim, M.Y., Ambesi-Impiombato, A., Perez-Garcia, A., Carpenter, Z., Abate, F., Allegretta, M., Haydu, J.E., Jiang, X., Lossos, I.S., Nicolas, C., Balbin, M., Bastard, C., Bhagat, G., Piris, M.A., Campo, E., Bernard, O.A., Rabadan, R. and Ferrando, A.A., 2014. Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nature Genetics*, 46(2), pp.166–170.

Pang, R., Law, W.L., Chu, A.C.Y., Poon, J.T., Lam, C.S.C., Chow, A.K.M., Ng, L., Cheung, L.W.H., Lan, X.R., Lan, H.Y., Tan, V.P.Y., Yau, T.C., Poon, R.T. and Wong, B.C.Y., 2010. A subpopulation of CD26 + cancer stem cells with metastatic capacity in human colorectal cancer. *Cell Stem Cell*, 6(6), pp.603–615.

Park, J.L., Kim, Hee Jin, Seo, E.H., Kwon, O.H., Lim, B., Kim, M., Kim, S.Y., Song, K.S., Kang, G.H., Kim, Hyun Ja, Choi, B.Y. and Kim, Y.S., 2015. Decrease of 5hmC in gastric cancers is associated with TET1 silencing due to with DNA methylation and bivalent histone marks at TET1 CpG island 3'-shore. *Oncotarget*, 6(35), pp.37647–37662.

Park, S., Han, S.S., Park, C.H., Hahm, E.R., Lee, S.J., Park, H.K., Lee, S.H., Kim, W.S.,

Jung, C.W., Park, K., Riordan, H.D., Kimler, B.F., Kim, K. and Lee, J.H., 2004. L-Ascorbic acid induces apoptosis in acute myeloid leukemia cells via hydrogen peroxide-mediated mechanisms. *International Journal of Biochemistry and Cell Biology*, 36(11), pp.2180–2195.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., Tahiliani, M., Daley, G.Q., Liu, X.S., Ecker, J.R., Milos, P.M., Agarwal, S. and Rao, A., 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, 473(7347), pp.394–397.

Patnaik, M.M., Hanson, C.A., Hodnefield, J.M., Lasho, T.L., Finke, C.M., Knudson, R.A., Ketterling, R.P., Pardanani, A. and Tefferi, A., 2012. Differential prognostic effect of IDH1 versus IDH2 mutations in myelodysplastic syndromes: A Mayo Clinic Study of 277 patients. *Leukemia*, 26(1), pp.101–105.

Pawlowska, E., Szczepanska, J. and Blasiak, J., 2019. Pro-and Antioxidant Effects of Vitamin C in Cancer in correspondence to Its Dietary and Pharmacological Concentrations. *Oxidative Medicine and Cellular Longevity*.

Pei, Y.F., Tao, R., Li, J.F., Su, L.P., Yu, B.Q., Wu, X.Y., Yan, M., Gu, Q.L., Zhu, Z.G. and Liu, B.Y., 2016. TET1 inhibits gastric cancer growth and metastasis by PTEN demethylation and re-expression. *Oncotarget*, 7(21), pp.31322–31335.

Peng, D., Ge, G., Gong, Y., Zhan, Y., He, S., Guan, B., Li, Y., Xu, Z., Hao, H., He, Z., Xiong, G., Zhang, C., Shi, Y., Zhou, Y., Ci, W., Li, X. and Zhou, L., 2018. Vitamin C increases 5-hydroxymethylcytosine level and inhibits the growth of bladder cancer. *Clinical Epigenetics*, 10(1), p.94.

Peng, W., Zhang, H., Tan, S., Li, Y., Zhou, Y., Wang, L., Liu, C., Li, Q., Cen, X., Yang, S. and Zhao, Y., 2020. Synergistic antitumor effect of 5-fluorouracil with the novel LSD1 inhibitor ZY0511 in colorectal cancer. *Therapeutic Advances in Medical Oncology*, 12.

Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K. and Yura, R., 1972. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *The Biochemical journal*, 126(4), pp.781–790.

Perera, A., Eisen, D., Wagner, M., Laube, S.K., Künzel, A.F., Koch, S., Steinbacher, J., Schulze, E., Splith, V., Mittermeier, N., Müller, M., Biel, M., Carell, T. and Michalakis, S., 2015. TET3 is recruited by REST for context-specific hydroxymethylation and induction of gene expression. *Cell Reports*, 11(2), pp.283–294.

Pfaffeneder, T., Hackner, B., Truß, M., Münzel, M., Müller, M., Deiml, C.A., Hagemeier, C. and Carell, T., 2011. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angewandte Chemie - International Edition*, 50(31), pp.7008–7012.

Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S.K., Eisen, D., Truss, M., Steinbacher, J., Hackner, B., Kotljarova, O., Schuermann, D., Michalakis, S., Kosmatchev, O., Schiesser, S., Steigenberger, B., Raddaoui, N., Kashiwazaki, G., Müller, U., Spruijt, C.G., Vermeulen, M., Leonhardt, H., Schär, P., Müller, M. and Carell, T., 2014. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nature Chemical Biology*, 10(7), pp.574–581.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research*, 29(9).

Pfeifer, G.P., Kadam, S. and Jin, S.G., 2013. 5-hydroxymethylcytosine and its potential roles in development and cancer. *Epigenetics and Chromatin*, 6(1), pp.1–9. BioMed Central.

Pitt, J.J., 2009. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *The Clinical biochemist. Reviews*, 30(1), pp.19–34.

Pivovarcikova, K., Agaimy, A., Martinek, P., Alaghehbandan, R., Perez-Montiel, D., Alvarado-Cabrero, I., Rogala, J., Kuroda, N., Rychly, B., Gasparov, S., Michalova, K., Michal, M., Hora, M., Pitra, T., Tuckova, I., Laciok, S., Mareckova, J. and Hes, O., 2019. Primary renal well-differentiated neuroendocrine tumour (carcinoid): next-generation sequencing study of 11 cases. *Histopathology*, 75(1), pp.104–117.

Polireddy, K., Dong, R., Reed, G., Yu, J., Chen, P., Williamson, S., Violet, P.C., Pessetto, Z., Godwin, A.K., Fan, F., Levine, M., Drisko, J.A. and Chen, Q., 2017. High dose parenteral ascorbate inhibited pancreatic cancer growth and metastasis: Mechanisms and a phase I/IIa study. *Scientific Reports*, 7(1), pp.1–15.

Proudhon, C., Duffié, R., Ajjan, S., Cowley, M., Iranzo, J., Carbajosa, G., Saadeh, H., Holland, M.L., Oakey, R.J., Rakyan, V.K., Schulz, R. and Bourc'his, D., 2012. Protection against De Novo Methylation Is Instrumental in Maintaining Parent-of-Origin Methylation Inherited from the Gametes. *Molecular Cell*, 47(6), pp.909–920.

Puig, I., Tenbaum, S.P., Chicote, I., Arqués, O., Martínez-Quintanilla, J., Cuesta-Borrás, E., Ramírez, L., Gonzalo, P., Soto, A., Aguilar, S., Eguizabal, C., Caratù, G., Prat, A., Argilés, G., Landolfi, S., Casanovas, O., Serra, V., Villanueva, A., Arroyo, A.G., Terracciano, L., Nuciforo, P., Seoane, J., Recio, J.A., Vivancos, A., Dienstmann, R., Tabernero, J. and Palmer, H.G., 2018. TET2 controls chemoresistant slow-cycling cancer cell survival and tumor recurrence. *Journal of Clinical Investigation*, 128(9), pp.3887–3905.

Qing, Y., Tian, Z., Bi, Y., Wang, Y., Long, J., Song, C.X. and Diao, J., 2017. Quantitation and mapping of the epigenetic marker 5-hydroxymethylcytosine. *BioEssays*, 39(5). John Wiley and Sons Inc.

Qiu, Z., Lin, A.-P., Jiang, S., Elkashef, S.M., Myers, J., Srikantan, S., Sasi, B., Cao, J.Z., Godley, L.A., Rakheja, D., Lyu, Y., Zheng, S., Madesh, M., Shiio, Y., Dahia, P.L.M. and Aguiar, R.C.T., 2020. MYC Regulation of D2HGDH and L2HGDH Influences the Epigenome and Epitranscriptome. *Cell Chemical Biology*.

Qu, G. zhi, Grundy, P.E., Narayan, A. and Ehrlich, M., 1999. Frequent hypomethylation in Wilms tumors of pericentromeric DNA in chromosomes 1 and 16. *Cancer Genetics and Cytogenetics*, 109(1), pp.34–39.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P. V., Grimaldi, G., Riccio, A. and Trono, D., 2011. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Molecular Cell*, 44(3), pp.361–372.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A. and Cairns, B.R., 2008. DNA Demethylation in Zebrafish Involves the Coupling of a Deaminase, a Glycosylase, and Gadd45. *Cell*, 135(7), pp.1201–1212.

Raiber, E.A., Murat, P., Chirgadze, D.Y., Beraldi, D., Luisi, B.F. and Balasubramanian,

S., 2015. 5-formylcytosine alters the structure of the DNA double helix. *Nature Structural and Molecular Biology*, 22(1), pp.44–49.

Ramsawhook, A., Lewis, L., Coyle, B. and Ruzov, A., 2017. Medulloblastoma and ependymoma cells display increased levels of 5-carboxylcytosine and elevated TET1 expression. *Clinical Epigenetics*, 9(1), p.18.

Randhawa, G.S., Cui, H., Barletta, J.A., Strichman-Almashanu, L.Z., Talpaz, M., Kantarjian, H., Deisseroth, A.B., Champlin, R.C. and Feinberg, A.P., 1998. Loss of imprinting in disease progression in chronic myelogenous leukemia. *Blood*, 91(9), pp.3144–3147.

Rao, V.K., Swarnaseetha, A., Tham, G.H., Lin, W.Q., Han, B. Bin, Benoukraf, T., Xu, G.L. and Ong, C.T., 2020. Phosphorylation of Tet3 by cdk5 is critical for robust activation of BRN2 during neuronal differentiation. *Nucleic acids research*, 48(3), pp.1225–1238.

Rawłuszko-Wieczorek, A.A., Siera, A., Horbacka, K., Horst, N., Krokowicz, P. and Jagodziński, P.P., 2015. Clinical significance of DNA methylation mRNA levels of TET family members in colorectal cancer. *Journal of Cancer Research and Clinical Oncology*, 141(8), pp.1379–1392.

Razin, A. and Riggs, A.D., 1980. DNA Methylation and gene function. *Science*, 210(4470), pp.604–610.

Reik, W., 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, 447(7143), pp.425–432. Nature Publishing Group.

Ren, A., Dong, Y., Tsoi, H. and Yu, J., 2015. Detection of miRNA as Non-Invasive Biomarkers of Colorectal Cancer. *International Journal of Molecular Sciences*, 16(2), pp.2810–2823.

Ricci-Vitiani, L., Lombardi, D.G., Pilozzi, E., Biffoni, M., Todaro, M., Peschle, C. and De Maria, R., 2007. Identification and expansion of human colon-cancer-initiating cells. *Nature*, 445(7123), pp.111–115.

Rodríguez-Aguilera, J.R., Ecsedi, S., Goldsmith, C., Cros, M.P., Domínguez-López, M., Guerrero-Celis, N., Pérez-Cabeza de Vaca, R., Chemin, I., Recillas-Targa, F., Chagoya de Sánchez, V. and Hernández-Vargas, H., 2020. Genome-wide 5-

hydroxymethylcytosine (5hmC) emerges at early stage of in vitro differentiation of a putative hepatocyte progenitor. *Scientific Reports*, 10(1).

Rodriguez, E.F., De Marchi, F., Lokhandwala, P.M., Belchis, D., Xian, R., Gocke, C.D., Eshleman, J.R., Illei, P. and Li, M., 2020. *IDH1* and *IDH2* mutations in lung adenocarcinomas: Evidences of subclonal evolution. *Cancer Medicine*, 9(12), pp.4386–4394.

Romero-Barrios, N., Legascue, M.F., Benhamed, M., Ariel, F. and Crespi, M., 2018. SURVEY AND SUMMARY Splicing regulation by long noncoding RNAs. *Nucleic Acids Research*, 46(5), pp.2169–2184.

Ross, S.E. and Bogdanovic, O., 2019. TET enzymes, DNA demethylation and pluripotency. *Biochemical Society Transactions*, 47(3), pp.875–885. Portland Press Ltd.

Rougier, N., Bourc'his, D., Molina Gomes, D., Niveleau, A., Plachot, M., Pàldi, A. and Viegas-Péquignot, E., 1998. Chromosome methylation patterns during mammalian preimplantation development. *Genes and Development*, 12(14), pp.2108–2113.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D.B. and Trono, D., 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, 463(7278), pp.237–240.

Sabry, D., El-Deek, S.E.M., Maher, M., El-Baz, M.A.H., El-Bader, H.M., Amer, E., Hassan, E.A., Fathy, W. and El-Deek, H.E.M., 2019. Role of miRNA-210, miRNA-21 and miRNA-126 as diagnostic biomarkers in colorectal carcinoma: impact of HIF-1α-VEGF signaling pathway. *Molecular and Cellular Biochemistry*, 454(1–2), pp.177–189.

Saeed, W.H., Eissa, A.A. and Al-Doski, A.A., 2019. Impact of TP53 gene promoter methylation on chronic lymphocytic leukemia pathogenesis and progression. *Journal of Blood Medicine*, 10, pp.399–404.

Saitou, M. and Yamaji, M., 2012. Primordial germ cells in mice. *Cold Spring Harbor Perspectives in Biology*, 4(11), p.a008375.

Sakai, T., Toguchida, J., Ohtani, N., Yandell, D.W., Rapaport, J.M. and Dryja, T.P.,

1991. Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *American Journal of Human Genetics*, 48(5), pp.880–888.

Sakata-Yanagimoto, M., Enami, T., Yoshida, K., Shiraishi, Y., Ishii, R., Miyake, Y., Muto, H., Tsuyama, N., Sato-Otsubo, A., Okuno, Y., Sakata, S., Kamada, Y., Nakamoto-Matsubara, R., Tran, N.B., Izutsu, K., Sato, Yusuke, Ohta, Y., Furuta, J., Shimizu, S., Komeno, T., Sato, Yuji, Ito, T., Noguchi, M., Noguchi, E., Sanada, M., Chiba, K., Tanaka, H., Suzukawa, K., Nanmoku, T., Hasegawa, Y., Nureki, O., Miyano, S., Nakamura, N., Takeuchi, K., Ogawa, S. and Chiba, S., 2014. Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nature Genetics*, 46(2), pp.171–175.

Sant, D.W., Mustafi, S., Gustafson, C.B., Chen, J., Slingerland, J.M. and Wang, G., 2018. Vitamin C promotes apoptosis in breast cancer cells by increasing TRAIL expression. *Scientific Reports*, 8(1), pp.1–11.

Sardina, J.L., Collombet, S., Tian, T. V., Gómez, A., Di Stefano, B., Berenguer, C., Brumbaugh, J., Stadhouders, R., Segura-Morales, C., Gut, M., Gut, I.G., Heath, S., Aranda, S., Di Croce, L., Hochedlinger, K., Thieffry, D. and Graf, T., 2018. Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. *Cell Stem Cell*, 23(5), pp.727-741.e9.

Sasaki, M., Knobbe, C.B., Itsumi, M., Elia, A.J., Harris, I.S., Chio, I.I.C., Cairns, R.A., Mccracken, S., Wakeham, A., Haight, J., Ten, A.Y., Snow, B., Ueda, T., Inoue, S., Yamamoto, K., Ko, M., Rao, A., Yen, K.E., Su, S.M. and Mak, T.W., 2012. D-2-hydroxyglutarate produced by mutant Idh1 perturbs collagen maturation and basement membrane function. *Genes and Development*, 26(18), pp.2038–2049.

Sato, Y., Yoshizato, T., Shiraishi, Y., Maekawa, S., Okuno, Y., Kamura, T., Shimamura, T., Sato-Otsubo, A., Nagae, G., Suzuki, H., Nagata, Y., Yoshida, K., Kon, A., Suzuki, Y., Chiba, K., Tanaka, H., Niida, A., Fujimoto, A., Tsunoda, T., Morikawa, T., Maeda, D., Kume, H., Sugano, S., Fukayama, M., Aburatani, H., Sanada, M., Miyano, S., Homma, Y. and Ogawa, S., 2013. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nature Genetics*, 45(8), pp.860–867.

Schenkel, L.C., Kernohan, K.D., McBride, A., Reina, D., Hodge, A., Ainsworth, P.J., Rodenhiser, D.I., Pare, G., Bérubé, N.G., Skinner, C., Boycott, K.M., Schwartz, C.

and Sadikovic, B., 2017. Identification of epigenetic signature associated with alpha thalassemia/mental retardation X-linked syndrome. *Epigenetics and Chromatin*, 10(1), pp.1–11.

Schenkel, L.C., Schwartz, C., Skinner, C., Rodenhiser, D.I., Ainsworth, P.J., Pare, G. and Sadikovic, B., 2016. Clinical Validation of Fragile X Syndrome Screening by DNA Methylation Array. *Journal of Molecular Diagnostics*, 18(6), pp.834–841.

Schnekenburger, M., Florean, C., Dicato, M. and Diederich, M., 2015. Epigenetic alterations as a universal feature of cancer hallmarks and a promising target for personalized treatments. *Current Topics in Medicinal Chemistry*, 16(7), pp.745–776.

Schomacher, L., Han, D., Musheev, M.U., Arab, K., Kienhöfer, S., Von Seggern, A. and Niehrs, C., 2016. Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nature Structural and Molecular Biology*, 23(2), pp.116–124.

Schønberg, S.A., Lundemo, A.G., Fladvad, T., Holmgren, K., Bremseth, H., Nilsen, A., Gederaas, O., Tvedt, K.E., Egeberg, K.W. and Krokan, H.E., 2006. Closely related colon cancer cell lines display different sensitivity to polyunsaturated fatty acids, accumulate different lipid classes and downregulate sterol regulatory element-binding protein 1. *FEBS Journal*, 273(12), pp.2749–2765.

Schubert, S.A., Morreau, H., de Miranda, N.F.C.C. and van Wezel, T., 2020. The missing heritability of familial colorectal cancer. *Mutagenesis*, 35(3), pp.221–231.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G. and Rauscher, F.J., 2002. SETDB1: A novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes and Development*, 16(8), pp.919–932.

Schwarz, B.A., Bar-Nur, O., Silva, J.C.R. and Hochedlinger, K., 2014. Nanog is dispensable for the generation of induced pluripotent stem cells. *Current Biology*, 24(3), pp.347–350.

Scourzic, L., Mouly, E. and Bernard, O.A., 2015. TET proteins and the control of cytosine demethylation in cancer. *Genome Medicine*, 7(1), p.9.

Secardin, L., Limia, C.E.G., di Stefano, A., Bonamino, M.H., Saliba, J., Kataoka, K., Rehen, S.K., Raslova, H., Marty, C., Ogawa, S., Vainchenker, W., Monte-Mor, B. da C.R. and Plo, I., 2020. TET2 haploinsufficiency alters reprogramming into induced pluripotent stem cells. *Stem Cell Research*, 44, p.101755.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W. and Reik, W., 2012. The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Molecular Cell*, 48(6), pp.849–862.

Seo, J.S., Choi, Y.H., Moon, J.W., Kim, H.S. and Park, S.H., 2017. Hinokitiol induces DNA demethylation via DNMT1 and UHRF1 inhibition in colon cancer cells. *BMC Cell Biology*, 18(1), pp.1–11.

Sérandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Palierne, G., Gheeraert, C., Barloy-Hubler, F., Péron, C. Le, Madigou, T., Durand, E., Froguel, P., Staels, B., Lefebvre, P., Métivier, R., Eeckhoute, J. and Salbert, G., 2012. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Research*, 40(17), pp.8255–8265.

Sheaffer, K.L., Elliott, E.N. and Kaestner, K.H., 2016. DNA hypomethylation contributes to genomic instability and intestinal cancer initiation. *Cancer Prevention Research*, 9(7), pp.534–546.

Sheaffer, K.L., Kim, R., Aoki, R., Elliott, E.N., Schug, J., Burger, L., Schübeler, D. and Kaestner, K.H., 2014. DNA methylation is required for the control of stem cell differentiation in the small intestine. *Genes and Development*, 28(6), pp.652–664.

Shen, L. and Zhang, Y., 2012. Enzymatic Analysis of Tet Proteins: Key Enzymes in the Metabolism of DNA Methylation. *Methods in Enzymology*. Academic Press Inc., pp.93–105.

Shen, Y., Zhu, Y.M., Fan, X., Shi, J.Y., Wang, Q.R., Yan, X.J., Gu, Z.H., Wang, Y.Y., Chen, B., Jiang, C.L., Yan, H., Chen, F.F., Chen, H.M., Chen, Z., Jin, J. and Chen, S.J., 2011. Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. *Blood*, 118(20), pp.5593–5603.

Shenoy, N., Bhagat, T., Nieves, E., Stenson, M., Lawson, J., Choudhary, G.S., Habermann, T., Nowakowski, G., Singh, R., Wu, X., Verma, A. and Witzig, T.E., 2017. Upregulation of TET activity with ascorbic acid induces epigenetic modulation of lymphoma cells. *Blood cancer journal*, 7(7), p.e587.

Shenoy, N., Bhagat, T.D., Cheville, J., Lohse, C., Bhattacharyya, S., Tischer, A., Machha, V., Gordon-Mitchell, S., Choudhary, G., Wong, L.F., Gross, L.A., Ressigue, E., Leibovich, B., Boorjian, S.A., Steidl, U., Wu, X., Pradhan, K., Gartrell, B., Agarwal, B., Pagliaro, L., Suzuki, M., Greally, J.M., Rakheja, D., Houston Thompson, R., Susztak, K., Witzig, T., Zou, Y. and Verma, A., 2019. Ascorbic acid–induced TET activation mitigates adverse hydroxymethylcytosine loss in renal cell carcinoma. *Journal of Clinical Investigation*, 129(4), pp.1612–1625.

Shi, D.Q., Ali, I., Tang, J. and Yang, W.C., 2017. New insights into 5hmC DNA modification: Generation, distribution and function. *Frontiers in Genetics*, 8(JUL), p.100. Frontiers Media S.A.

Shi, F.T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M.A., Wan, M. and Songyang, Z., 2013. Ten-eleven translocation 1 (Tet1) is regulated by o-linked n-acetylglucosamine transferase (ogt) for target gene repression in mouse embryonic stem cells. *Journal of Biological Chemistry*, 288(29), pp.20776–20784.

Shi, M., Zhang, H., Wang, L., Zhu, C., Sheng, K., Du, Y., Wang, K., Dias, A., Chen, S., Whitman, M., Wang, E., Reed, R. and Cheng, H., 2015. Premature termination codons are recognized in the nucleus in a reading-frame-dependent manner. *Cell Discovery*, 1(1), pp.1–20.

Shibata, T., Kokubu, A., Miyamoto, M., Sasajima, Y. and Yamazaki, N., 2011. Mutant IDH1 confers an in vivo growth in a melanoma cell line with BRAF mutation. *American Journal of Pathology*, 178(3), pp.1395–1402.

Shih, A.H., Abdel-Wahab, O., Patel, J.P. and Levine, R.L., 2012. The role of mutations in epigenetic regulators in myeloid malignancies. *Nature Reviews Cancer*, 12(9), pp.599–612. Nature Publishing Group.

Shrestha, R., Sakata-Yanagimoto, M., Maie, K., Oshima, M., Ishihara, M., Suehara, Y., Fukumoto, K., Nakajima-Takagi, Y., Matsui, H., Kato, T., Muto, H., Sakamoto, T.,

Kusakabe, M., Nannya, Y., Makishima, H., Ueno, H., Saiki, R., Ogawa, S., Chiba, K., Shiraishi, Y., Miyano, S., Mouly, E., Bernard, O.A., Inaba, T., Koseki, H., Iwama, A. and Chiba, S., 2020. Molecular pathogenesis of progression to myeloid leukemia from TET-insufficient status. *Blood Advances*, 4(5), pp.845–854.

Siegel, R.L., Miller, K.D., Goding Sauer, A., Fedewa, S.A., Butterly, L.F., Anderson, J.C., Cercek, A., Smith, R.A. and Jemal, A., 2020. Colorectal cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(3), pp.145–164.

Singh, S.K., Hawkins, C., Clarke, I.D., Squire, J.A., Bayani, J., Hide, T., Henkelman, R.M., Cusimano, M.D. and Dirks, P.B., 2004. Identification of human brain tumour initiating cells. *Nature*, 432(7015), pp.396–401.

Sirico, M.L., Guida, B., Procino, A., Pota, A., Sodo, M., Grandaliano, G., Simone, S., Pertosa, G., Riccio, E. and Memoli, B., 2012. Human mature adipocytes express albumin and this expression is not regulated by inflammation. *Mediators of Inflammation*, 2012, pp.236796–236796.

Siuzdak, G., 2004. An introduction to mass spectrometry ionization: An excerpt from The Expanding Role of Mass Spectrometry in Biotechnology, 2nd ed.; MCC Press: San Diego, 2005. *Journal of the Association for Laboratory Automation*, 9(2), pp.50–63.

Slater, C., De La Mare, J.A. and Edkins, A.L., 2018. In vitro analysis of putative cancer stem cell populations and chemosensitivity in the SW480 and SW620 colon cancer metastasis model. *Oncology Letters*, 15(6), pp.8516–8526.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A. and Meissner, A., 2012. A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, 484(7394), pp.339–344.

Smits, A.H., Ziebell, F., Joberty, G., Zinn, N., Mueller, W.F., Clauder-Münster, S., Eberhard, D., Fälth Savitski, M., Grandi, P., Jakob, P., Michon, A.M., Sun, H., Tessmer, K., Bürckstümmer, T., Bantscheff, M., Steinmetz, L.M., Drewes, G. and Huber, W., 2019. Biological plasticity rescues target activity in CRISPR knock outs. *Nature Methods*, 16(11), pp.1087–1093.

Song, S.J., Ito, K., Ala, U., Kats, L., Webster, K., Sun, S.M., Jongen-Lavrencic, M.,

Manova-Todorova, K., Teruya-Feldstein, J., Avigan, D.E., Delwel, R. and Pandolfi, P.P., 2013. The oncogenic MicroRNA miR-22 targets the TET2 tumor suppressor to promote hematopoietic stem cell self-renewal and transformation. *Cell Stem Cell*, 13(1), pp.87–101.

Souren, N.Y., Gerdes, L.A., Lutsik, P., Gasparoni, G., Beltrán, E., Salhab, A., Kümpfel, T., Weichenhan, D., Plass, C., Hohlfeld, R. and Walter, J., 2019. DNA methylation signatures of monozygotic twins clinically discordant for multiple sclerosis. *Nature Communications*, 10(1), pp.1–12.

De Sousa E Melo, F., Kurtova, A. V., Harnoss, J.M., Kljavin, N., Hoeck, J.D., Hung, J., Anderson, J.E., Storm, E.E., Modrusan, Z., Koeppen, H., Dijkgraaf, G.J.P., Piskol, R. and De Sauvage, F.J., 2017. A distinct role for Lgr5 + stem cells in primary and metastatic colon cancer. *Nature*, 543(7647), pp.676–680.

De Souza, A., Tinguely, M., Pfaltz, M., Burghart, D.R. and Kempf, W., 2014. Loss of expression of 5-hydroxymethylcytosine in CD30-positive cutaneous lymphoproliferative disorders. *Journal of Cutaneous Pathology*, 41(12), pp.901–906.

Soverini, S., Score, J., Iacobucci, I., Poerio, A., Lonetti, A., Gnani, A., Colarossi, S., Ferrari, A., Castagnetti, F., Rosti, G., Cervantes, F., Hochhaus, A., Delledonne, M., Ferrarini, A., Sazzini, M., Luiselli, D., Baccarani, M., Cross, N.C.P. and Martinelli, G., 2011. IDH2 somatic mutations in chronic myeloid leukemia patients in blast crisis. *Leukemia*, 25(1), pp.178–181. Nature Publishing Group.

Spans, L., Van den Broeck, T., Smeets, E., Prekovic, S., Thienpont, B., Lambrechts, D., Jeffrey Karnes, R., Erho, N., Alshalalfa, M., Davicioni, E., Helsen, C., Gevaert, T., Tosco, L., Haustermans, K., Lerut, E., Joniau, S. and Claessens, F., 2016. Genomic and epigenomic analysis of high-risk prostate cancer reveals changes in hydroxymethylation and TET1. *Oncotarget*, 7(17), pp.24326–24338.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H.C., Mensinga, A., Brinkman, A.B., Lephikov, K., Müller, U., Walter, J., Boelens, R., Van Ingen, H., Leonhardt, H., Carell, T. and Vermeulen, M., 2013. Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. *Cell*, 152(5), pp.1146–1159.

Stadtfeld, M., Apostolou, E., Ferrari, F., Choi, J., Walsh, R.M., Chen, T., Ooi, S.S.K., Kim, S.Y., Bestor, T.H., Shioda, T., Park, P.J. and Hochedlinger, K., 2012. Ascorbic acid prevents loss of Dlk1-Dio3 imprinting and facilitates generation of allg-iPS cell mice from terminally differentiated B cells. *Nature Genetics*, 44(4), pp.398–405.

Stahl, M., Kohrman, N., Gore, S.D., Kim, T.K., Zeidan, A.M. and Prebet, T., 2016. Epigenetics in Cancer: A Hematological Perspective. *PLoS Genetics*, 12(10), p.e1006193. Public Library of Science.

Steenman, M.J.C., Rainier, S., Dobry, C.J., Grundy, P., Horon, I.L. and Feinberg, A.P., 1994. Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. *Nature Genetics*, 7(3), pp.433–439.

Stoffel, E.M. and Kastrinos, F., 2014. Familial colorectal cancer, beyond lynch syndrome. *Clinical Gastroenterology and Hepatology*, 12(7), pp.1059–1068.

Stoll, G.A., Oda, S. ichiro, Chong, Z.S., Yu, M., McLaughlin, S.H. and Modis, Y., 2019. Structure of KAP1 tripartite motif identifies molecular interfaces required for retroelement silencing. *Proceedings of the National Academy of Sciences of the United States of America*, 116(30), pp.15042–15051.

Storebjerg, T.M., Strand, S.H., Høyer, S., Lynnerup, A.S., Borre, M., Ørntoft, T.F. and Sørensen, K.D., 2018. Dysregulation and prognostic potential of 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) levels in prostate cancer. *Clinical Epigenetics*, 10(1), pp.1–16.

Stricker, S.H. and Götz, M., 2018. DNA-methylation: Master or slave of neural fate decisions? *Frontiers in Neuroscience*, 12(FEB), p.5. Frontiers Media S.A.

Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. and Jacobsen, S.E., 2011. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biology*, 12(6), pp.1–8.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H. and Tajima, S., 2004. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *Journal of Biological Chemistry*, 279(26), pp.27816–27823.

Suhara, T., Hishiki, T., Kasahara, M., Hayakawa, N., Oyaizu, T., Nakanishi, T., Kubo, A., Morisaki, H., Kaelin, W.G., Suematsu, M. and Minamishima, Y.A., 2015. Inhibition of the oxygen sensor PHD2 in the liver improves survival in lactic acidosis by activating the Cori cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 112(37), pp.11642–11647.

Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. and Leonhardt, H., 2010. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Research*, 38(19), pp.e181–e181.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. and Rao, A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324(5929), pp.930–935.

Takahashi, K. and Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4), pp.663–676.

Tang, G.Y., Tang, G.J., Yin, L., Chao, C., Zhou, R., Ren, G.P., Chen, J.Y. and Zhang, W., 2019. ECRG4 acts as a tumor suppressor gene frequently hypermethylated in human breast cancer. *Bioscience Reports*, 39(5).

Tate, P.H. and Bird, A.P., 1993. Effects of DNA methylation on DNA-binding proteins and gene expression. *Current Opinion in Genetics and Development*, 3(2), pp.226–231.

Tefferi, A., Lim, K.H., Abdel-Wahab, O., Lasho, T.L., Patel, J., Patnaik, M.M., Hanson, C.A., Pardanani, A., Gilliland, D.G. and Levine, R.L., 2009. Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML. *Leukemia*, 23(7), pp.1343–1345. Nature Publishing Group.

Teng, S., Ma, C., Yu, Y. and Yi, C., 2019. Hydroxyurea promotes TET1 expression and induces apoptosis in osteosarcoma cells. *Bioscience Reports*, 39(5).

Thienpont, B., Steinbacher, J., Zhao, H., D'Anna, F., Kuchnio, A., Ploumakis, A., Ghesquière, B., Van Dyck, L., Boeckx, B., Schoonjans, L., Hermans, E., Amant,

F., Kristensen, V.N., Koh, K.P., Mazzone, M., Coleman, M.L., Carell, T., Carmeliet, P. and Lambrechts, Di., 2016. Tumour hypoxia causes DNA hypermethylation by reducing TET activity. *Nature*, 537(7618), pp.63–68.

Thomson, J.P., Ottaviano, R., Unterberger, E.B., Lempiainen, H., Muller, A., Terranova, R., Illingworth, R.S., Webb, S., Kerr, A.R.W., Lyall, M.J., Drake, A.J., Rolandwolf, C., Moggs, J.G., Schwarz, M. and Meehan, R.R., 2016. Loss of tet1-associated5-hydroxymethylcytosine is concomitant with aberrant promoter hypermethylation in liver cancer. *Cancer Research*, 76(10), pp.3097–3108.

Tian, Y., Pan, F., Sun, X., Gan, M., Lin, A., Zhang, D., Zhu, Y. and Lai, M., 2017. Association of TET1 expression with colorectal cancer progression. *Scandinavian Journal of Gastroenterology*, 52(3), pp.312–320.

Tian, Y. ping, Lin, A. fen, Gan, M. fu, Wang, H., Yu, D., Lai, C., Zhang, D. dan, Zhu, Y. min and Lai, M. de, 2017. Global changes of 5-hydroxymethylcytosine and 5-methylcytosine from normal to tumor tissues are associated with carcinogenesis and prognosis in colorectal cancer. *Journal of Zhejiang University: Science B*, 18(9), pp.747–756.

Todaro, M., Gaggianesi, M., Catalano, V., Benfante, A., Iovino, F., Biffoni, M., Apuzzo, T., Sperduti, I., Volpe, S., Cocorullo, G., Gulotta, G., Dieli, F., De Maria, R. and Stassi, G., 2014. CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. *Cell Stem Cell*, 14(3), pp.342–356.

Tomita, N., Jiang, W., Hibshoosh, H., Warburton, D., Kahn, S.M., Weinstein, I.B. and Development, [D W, 1992. Isolation and Characterization of a Highly Malignant Variant of the SW480 Human Colon Cancer Cell Line. *Cancer Research*, 52, pp.6840–6847.

Tong, M., Gao, S., Qi, W., Shi, C., Qiu, M., Yang, F., Bai, S., Li, H., Wang, Z., Sun, Z., Wang, L. and Che, Y., 2019. 5-hydroxymethylcytosine as a potential epigenetic biomarker in papillary thyroid carcinoma. *Oncology Letters*, 18(3), pp.2304–2309.

Towers, A.J., Tremblay, M.W., Chung, L., Li, X.L., Bey, A.L., Zhang, W., Cao, X., Wang, X., Wang, P., Duffney, L.J., Siecinski, S.K., Xu, S., Kim, Y., Kong, X., Gregory, S., Xie, W. and Jiang, Y.H., 2018. Epigenetic dysregulation of Oxtr in

Tet1-deficient mice has implications for neuropsychiatric disorders. *JCI insight*, 3(23).

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B. and Issa, J.P.J., 1999. CpG island methylator phenotype in colorectal cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 96(15), pp.8681–8686.

Tsagaratou, A., Äijö, T., Lio, C.W.J., Yue, X., Huang, Y., Jacobsen, S.E., Lähdesmäki, H. and Rao, A., 2014. Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 111(32), pp.E3306–E3315.

Tsai, K.W., Li, G.C., Chen, C.H., Yeh, M.H., Huang, J.S., Tseng, H.H., Fu, T.Y., Liou, H.H., Pan, H.W., Huang, S.F., Chen, C.C., Chang, H.Y., Ger, L.P. and Chang, H.T., 2015. Reduction of global 5-hydroxymethylcytosine is a poor prognostic factor in breast cancer patients, especially for an ER/PR-negative subtype. *Breast Cancer Research and Treatment*, 153(1), pp.219–234.

Tu, J., Ng, S.H., Shui Luk, A.C., Liao, J., Jiang, X., Feng, B., Lun Mak, K.K., Rennert, O.M., Chan, W.Y. and Lee, T.L., 2015. MicroRNA-29b/Tet1 regulatory axis epigenetically modulates mesendoderm differentiation in mouse embryonic stem cells. *Nucleic Acids Research*, 43(16), pp.7805–7822.

Tucker, D.W., Getchell, C.R., McCarthy, E.T., Ohman, A.W., Sasamoto, N., Xu, S., Ko, J.Y., Gupta, M., Shafrir, A., Medina, J.E., Lee, J.J., MacDonald, L.A., Malik, A., Hasselblatt, K.T., Li, W., Zhang, H., Kaplan, S.J., Murphy, G.F., Hirsch, M.S., Liu, J.F., Matulonis, U.A., Terry, K.L., Lian, C.G. and Dinulescu, D.M., 2018. Epigenetic reprogramming strategies to reverse global loss of 5-hydroxymethylcytosine, a prognostic factor for poor survival in high-grade serous ovarian cancer. *Clinical Cancer Research*, 24(6), pp.1389–1401.

Tuladhar, R., Yeu, Y., Tyler Piazza, J., Tan, Z., Rene Clemenceau, J., Wu, X., Barrett, Q., Herbert, J., Mathews, D.H., Kim, J., Hyun Hwang, T. and Lum, L., 2019. CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nature Communications*, 10(1), pp.1–10.

Uetaki, M., Tabata, S., Nakasuka, F., Soga, T. and Tomita, M., 2015. Metabolomic alterations in human cancer cells by Vitamin C-induced oxidative stress. *Scientific Reports*, 5(1), pp.1–9.

Uribe-Lewis, S., Carroll, T., Menon, S., Nicholson, A., Manasterski, P.J., Winton, D.J., Buczacki, S.J.A. and Murrell, A., 2020. 5-hydroxymethylcytosine and gene activity in mouse intestinal differentiation. *Scientific Reports*, 10(1), pp.1–11.

Uribe-Lewis, S., Stark, R., Carroll, T., Dunning, M.J., Bachman, M., Ito, Y., Stojic, L., Halim, S., Vowler, S.L., Lynch, A.G., Delatte, B., de Bony, E.J., Colin, L., Defrance, M., Krueger, F., Silva, A.L., ten Hoopen, R., Ibrahim, A.E.K., Fuks, F. and Murrell, A., 2015. 5-hydroxymethylcytosine marks promoters in colon that resist DNA hypermethylation in cancer. *Genome Biology*, 16(1), p.69.

Valderrama-Treviño, A.I., Barrera-Mera, B., Ceballos-Villalva, J.C. and Montalvo-Javé, E.E., 2017. Hepatic Metastasis from Colorectal Cancer. *Euroasian Journal of Hepato-Gastroenterology*, 7(2), pp.166–175.

Valinluck, V. and Sowers, L.C., 2007. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Research*, 67(3), pp.946–950.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, 3(7), p.research0034.1.

Vatandoust, S., Price, T.J. and Karapetis, C.S., 2015. Colorectal cancer: Metastases to a single organ. *World Journal of Gastroenterology*, 21(41), pp.11767–11776.

Vella, P., Scelfo, A., Jammula, S.G., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K. and Pasini, D., 2013. Tet Proteins Connect the O-Linked N-acetylglucosamine Transferase Ogt to Chromatin in Embryonic Stem Cells. *Molecular Cell*, 49(4), pp.645–656.

Verma, N., Pan, H., Doré, L.C., Shukla, A., Li, Q. V., Pelham-Webb, B., Teijeiro, V., González, F., Krivtsov, A., Chang, C.J., Papapetrou, E.P., He, C., Elemento, O. and Huangfu, D., 2018. TET proteins safeguard bivalent promoters from de novo

methylation in human embryonic stem cells. *Nature Genetics*, 50(1), pp.83–95.

Villena, J.A., Hock, M.B., Chang, W.Y., Barcas, J.E., Giguère, V. and Kralli, A., 2007. Orphan nuclear receptor estrogen-related receptor α is essential for adaptive themogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 104(4), pp.1418–1423.

Vincent, J.J., Huang, Y., Chen, P.Y., Feng, S., Calvopiña, J.H., Nee, K., Lee, S.A., Le, T., Yoon, A.J., Faull, K., Fan, G., Rao, A., Jacobsen, S.E., Pellegrini, M. and Clark, A.T., 2013. Stage-specific roles for Tet1 and Tet2 in DNA demethylation in primordial germ cells. *Cell Stem Cell*, 12(4), pp.470–478.

Visvader, J.E., 2011. Cells of origin in cancer. *Nature*, 469(7330), pp.314–322. Nature Publishing Group.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A. and Kinzler, K.W., 2013. Cancer genome landscapes. *Science*, 340(6127), pp.1546–1558. American Association for the Advancement of Science.

Vuletic, I., Zhou, K., Li, H., Bai, H., Meng, X., Zhu, S., Ding, Y., Li, J., Sun, H. and Ren, Q., 2017. Validation of Bevacizumab Therapy Effect on Colon Cancer Subtypes by Using Whole Body Imaging in Mice. *Molecular Imaging and Biology*, 19(6), pp.847–856.

Vymetalkova, V., Vodicka, P., Vodenkova, S., Alonso, S. and Schneider-Stock, R., 2019. DNA methylation and chromatin modifiers in colorectal cancer. *Molecular Aspects of Medicine*, 69, pp.73–92. Elsevier Ltd.

Waddington, C.H., 2012. The epigenotype. 1942. *International journal of epidemiology*, 41(1), pp.10–13.

Wang, J., Hevi, S., Kurash, J.K., Lei, H., Gay, F., Bajko, J., Su, H., Sun, W., Chang, H., Xu, G., Gaudet, F., Li, E. and Chen, T., 2009. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. *Nature Genetics*, 41(1), pp.125–129.

Wang, K., Liu, W., Yan, X.L., Li, J. and Xing, B.C., 2017. Long-term postoperative survival prediction in patients with colorectal liver metastasis. *Oncotarget*, 8(45), pp.79927–79934.

Wang, L., Ozark, P.A., Smith, E.R., Zhao, Z., Marshall, S.A., Rendleman, E.J., Piunti, A., Ryan, C., Whelan, A.L., Helmin, K.A., Morgan, M.A., Zou, L., Singer, B.D. and Shilatifard, A., 2018. TET2 coactivates gene expression through demethylation of enhancers. *Science Advances*, 4(11), p.eaau6986.

Wang, L., Zhang, Jun, Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, Jing, Li, G., Ci, W., Li, W., Zhou, Q., Aluru, N., Tang, F., He, C., Huang, X. and Liu, J., 2014. Programming and inheritance of parental DNA methylomes in mammals. *Cell*, 157(4), pp.979–991.

Wang, L., Zhou, Y., Xu, L., Xiao, R., Lu, X., Chen, L., Chong, J., Li, H., He, C., Fu, X.D. and Wang, D., 2015. Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature*, 523(7562), pp.621–625.

Wang, T., Pan, Q., Lin, L., Szulwach, K.E., Song, C.X., He, C., Wu, H., Warren, S.T., Jin, P., Duan, R. and Li, X., 2012. Genome-wide DNA hydroxymethylation changes are associated with neurodevelopmental genes in the developing human cerebellum. *Human Molecular Genetics*, 21(26), pp.5500–5510.

Wang, Yu, Alhaque, S., Cameron, K., Meseguer-Ripolles, J., Lucendo-Villarin, B., Rashidi, H. and Hay, D.C., 2017. Defined and scalable generation of hepatocyte-like cells from human pluripotent stem cells. *Journal of Visualized Experiments*, 2017(121), p.55355.

Wang, Y. and Zhang, Y., 2014. Regulation of TET protein stability by calpains. *Cell Reports*, 6(2), pp.278–284.

Wang, Yi, Hu, H., Wang, Q., Li, Z., Zhu, Y., Zhang, W., Wang, Yanling, Jiang, H. and Cheng, J., 2017. The level and clinical significance of 5-hydroxymethylcytosine in oral squamous cell carcinoma: An immunohistochemical study in 95 patients. *Pathology Research and Practice*, 213(8), pp.969–974.

Waters, 2019. *Xevo TqD Instrument specifications* [Online]. Available from: https://www.waters.com/webassets/cms/library/docs/720003973en.pdf [Accessed 1 May 2020].

Weber, A.R., Krawczyk, C., Robertson, A.B., Kusnierczyk, A., Vågbø, C.B., Schuermann, D., Klungland, A. and Schar, P., 2016. Biochemical reconstitution of

TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nature Communications*, 7(1), pp.1–13.

Weissmann, S., Alpermann, T., Grossmann, V., Kowarsch, A., Nadarajah, N., Eder, C., Dicker, F., Fasan, A., Haferlach, C., Haferlach, T., Kern, W., Schnittger, S. and Kohlmann, A., 2012. Landscape of TET2 mutations in acute myeloid leukemia. *Leukemia*, 26(5), pp.934–942.

Weksberg, R., Nishikawa, J., Caluseriu, O., Fei, Y.L., Shuman, C., Wei, C., Steele, L., Cameron, J., Smith, A., Ambus, I., Li, M., Ray, P.N., Sadowski, P. and Squire, J., 2001. Tumor development in the Beckwith-Wiedemann syndrome is associated with a variety of constitutional molecular 11p15 alterations including imprinting defects of KCNQ1OT1. *Human Molecular Genetics*, 10(26), pp.2989–3000.

Wernig-Zorc, S., Yadav, M.P., Kopparapu, P.K., Bemark, M., Kristjansdottir, H.L., Andersson, P.O., Kanduri, C. and Kanduri, M., 2019. Global distribution of DNA hydroxymethylation and DNA methylation in chronic lymphocytic leukemia. *Epigenetics and Chromatin*, 12(1), p.4.

Wilm, M., 2011. Principles of electrospray ionization. *Molecular and Cellular Proteomics*, 10(7). American Society for Biochemistry and Molecular Biology.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W. and Walter, J., 2011. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature Communications*, 2(1), pp.1–8.

Wu, D., Hu, D., Chen, H., Shi, G., Fetahu, I.S., Wu, F., Rabidou, K., Fang, R., Tan, L., Xu, S., Liu, H., Argueta, C., Zhang, L., Mao, F., Yan, G., Chen, J., Dong, Z., Lv, R., Xu, Y., Wang, M., Ye, Y., Zhang, S., Duquette, D., Geng, S., Yin, C., Lian, C.G., Murphy, G.F., Adler, G.K., Garg, R., Lynch, L., Yang, P., Li, Y., Lan, F., Fan, J., Shi, Y. and Shi, Y.G., 2018. Glucose-regulated phosphorylation of TET2 by AMPK reveals a pathway linking diabetes to cancer. *Nature*, 559(7715), pp.637–641.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Eve Sun, Y. and Zhang, Y., 2011. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature*, 473(7347), pp.389–394.

Wu, H., Zhao, M., Tan, L. and Lu, Q., 2016. The key culprit in the pathogenesis of systemic lupus erythematosus: Aberrant DNA methylation. *Autoimmunity Reviews*, 15(7), pp.684–689. Elsevier B.V.

Wu, J., Li, H., Shi, M., Zhu, Y., Ma, Y., Zhong, Y., Xiong, C., Chen, H. and Peng, C., 2019. TET1-mediated DNA hydroxymethylation activates inhibitors of the Wnt/β-catenin signaling pathway to suppress EMT in pancreatic tumor cells. *Journal of Experimental and Clinical Cancer Research*, 38(1), p.348.

Wu, M., Zhang, Y., Tang, A. and Tian, L., 2016. miR-506 inhibits cell proliferation and invasion by targeting TET family in colorectal cancer. *Iranian journal of basic medical sciences*, 19(3), pp.316–22.

Wu, M.Z., Chen, S.F., Nieh, S., Benner, C., Ger, L.P., Jan, C.I., Ma, L., Chen, C.H., Hishida, T., Chang, H.T., Lin, Y.S., Montserrat, N., Gascon, P., Sancho-Martinez, I. and Belmonte, J.C.I., 2015. Hypoxia drives breast tumor malignancy through a TET-TNFα-p38-MAPK signaling axis. *Cancer Research*, 75(18), pp.3912–3924.

Wu, X. and Zhang, Y., 2017. TET-mediated active DNA demethylation: Mechanism, function and beyond. *Nature Reviews Genetics*, 18(9), pp.517–534. Nature Publishing Group.

Wyatt, G.R. and Cohen, S.S., 1952. A new pyrimidine base from bacteriophage nucleic acids [7]. *Nature*, 170(4338), pp.1072–1073. Nature.

Xiao, M., Yang, H., Xu, W., Ma, S., Lin, H., Zhu, H., Liu, L., Liu, Y., Yang, C., Xu, Y., Zhao, S., Ye, D., Xiong, Y. and Guan, K.L., 2012. Inhibition of α-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes and Development*, 26(12), pp.1326–1338.

Xie, M., Lu, C., Wang, J., McLellan, M.D., Johnson, K.J., Wendl, M.C., McMichael, J.F., Schmidt, H.K., Yellapantula, V., Miller, C.A., Ozenberger, B.A., Welch, J.S., Link, D.C., Walter, M.J., Mardis, E.R., Dipersio, J.F., Chen, F., Wilson, R.K., Ley, T.J. and Ding, L., 2014. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature Medicine*, 20(12), pp.1472–1478.

Xie, Y.H., Chen, Y.X. and Fang, J.Y., 2020. Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduction and Targeted Therapy*, 5(1). Springer Nature.

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, Ping, Kim, S.H., Ito, S., Yang, C., Wang, Pu, Xiao, M.T., Liu, L.X., Jiang, W.Q., Liu, J., Zhang, J.Y., Wang, B., Frye, S., Zhang, Y., Xu, Y.H., Lei, Q.Y., Guan, K.L., Zhao, S.M. and Xiong, Y., 2011. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer Cell*, 19(1), pp.17–30.

Xu, X., Tan, X., Tampe, B., Wilhelmi, T., Hulshoff, M.S., Saito, S., Moser, T., Kalluri, R., Hasenfuss, G., Zeisberg, E.M. and Zeisberg, M., 2018. High-fidelity CRISPR/Cas9- based gene-specific hydroxymethylation rescues gene expression and attenuates renal fibrosis. *Nature Communications*, 9(1).

Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G.L. and Hu, R., 2016. A CRISPR-based approach for targeted DNA demethylation. *Cell Discovery*, 2(1), pp.1–12.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., Min, J., Nicholson, T., Chen, T., Xu, G., Shi, Y., Zhang, K. and Shi, Y.G., 2011. Genome-wide Regulation of 5hmC, 5mC, and Gene Expression by Tet1 Hydroxylase in Mouse Embryonic Stem Cells. *Molecular Cell*, 42(4), pp.451–464.

Xu, Y., Zhang, L., Wang, Q. and Zheng, M., 2020. Comparison of Different Colorectal Cancer With Liver Metastases Models Using Six Colorectal Cancer Cell Lines. *Pathology and Oncology Research*, pp.1–7.

Xu, Y.P., Lv, L., Liu, Y., Smith, M.D., Li, W.C., Tan, X.M., Cheng, M., Li, Z., Bovino, M., Aubé, J. and Xiong, Y., 2019. Tumor suppressor TET2 promotes cancer immunity and immunotherapy efficacy. *Journal of Clinical Investigation*, 129(10), pp.4316–4331.

Yamaguchi, S., Hong, K., Liu, R., Inoue, A., Shen, L., Zhang, K. and Zhang, Y., 2013. Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Research*, 23(3), pp.329–339.

Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K. and Saitou, M., 2013. PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell*, 12(3), pp.368–382.

Yan, H., Parsons, D.W., Jin, G., McLendon, R., Rasheed, B.A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., Riggins, G.J., Friedman, H., Friedman, A., Reardon, D., Herndon, J., Kinzler, K.W., Velculescu, V.E., Vogelstein, B. and Bigner, D.D., 2009. *IDH1* and *IDH2* Mutations in Gliomas. *New England Journal of Medicine*, 360(8), pp.765–773.

Yang, C., Ota-Kurogi, N., Ikeda, K., Okumura, T., Horie-Inoue, K., Takeda, S. and Inoue, S., 2020. MicroRNA-191 regulates endometrial cancer cell growth via TET1-mediated epigenetic modulation of APC. *Journal of biochemistry*, 168(1), pp.7–14.

Yang, J., Horton, J.R., Li, J., Huang, Y., Zhang, X., Blumenthal, R.M. and Cheng, X., 2019. Structural basis for preferential binding of human TCF4 to DNA containing 5-carboxylcytosine. *Nucleic acids research*, 47(16), pp.8375–8387.

Yang, Q., Liang, X., Sun, X., Zhang, L., Fu, X., Rogers, C.J., Berim, A., Zhang, S., Wang, S., Wang, B., Foretz, M., Viollet, B., Gang, D.R., Rodgers, B.D., Zhu, M.J. and Du, M., 2016. AMPK/α-Ketoglutarate Axis Dynamically Mediates DNA Demethylation in the Prdm16 Promoter and Brown Adipogenesis. *Cell Metabolism*, 24(4), pp.542–554.

Yang, Q., Wu, K., Ji, M., Jin, W., He, N., Shi, B. and Hou, P., 2013. Decreased 5-hydroxymethylcytosine (5-hmC) is an independent poor prognostic factor in gastric cancer patients. *Journal of Biomedical Nanotechnology*, 9(9), pp.1607–1616.

Yang, X., Liu, M., Li, M., Zhang, S., Hiju, H., Sun, J., Mao, Z., Zheng, M. and Feng, B., 2020. Epigenetic modulations of noncoding RNA: A novel dimension of Cancer biology. *Molecular Cancer*, 19(1), pp.1–12. BioMed Central Ltd.

Yang, Y.A., Zhao, J.C., Fong, K.W., Kim, J., Li, S., Song, C., Song, B., Zheng, B., He, C. and Yu, J., 2016. FOXA1 potentiates lineage-specific enhancer activation through modulating TET1 expression and function. *Nucleic Acids Research*,

44(17), pp.8153–8164.

Ye, Z., Li, J., Han, X., Hou, H., Chen, H., Zheng, X., Lu, J., Wang, L., Chen, W., Li, X. and Zhao, L., 2016. TET3 inhibits TGF-β1-induced epithelial-mesenchymal transition by demethylating miR-30d precursor gene in ovarian cancer cells. *Journal of Experimental and Clinical Cancer Research*, 35(1), p.72.

Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J. and Fazzio, T.G., 2011. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell*, 147(7), pp.1498–1510.

Yin, R., Mo, J., Dai, J. and Wang, H., 2017. Nickel(II) Inhibits Tet-Mediated 5-Methylcytosine Oxidation by High Affinity Displacement of the Cofactor Iron(II). *ACS Chemical Biology*, 12(6), pp.1494–1498.

Yin, R., Mo, J., Dai, J. and Wang, H., 2018. Nickel(II) inhibits the oxidation of DNA 5-methylcytosine in mammalian somatic cells and embryonic stem cells. *Metallomics*, 10(3), pp.504–512.

Yin, W., Wang, X., Li, Y., Wang, B., Song, M., Hulbert, A., Chen, C. and Yu, F., 2020. Promoter hypermethylation of cysteine dioxygenase type 1 in patients with non-small cell lung cancer. *Oncology Letters*, 20(1), pp.967–973.

Yin, X. and Xu, Y., 2016. Structure and function of TET enzymes. *Advances in Experimental Medicine and Biology*, 945, pp.275–302.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., Min, J.H., Jin, P., Ren, B. and He, C., 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*, 149(6), pp.1368–1380.

Yu, S., Yin, Y., Hong, S., Cao, S., Huang, Y., Chen, S., Liu, Y., Guan, H., Zhang, Q., Li, Y. and Xiao, H., 2020. TET1 is a Tumor Suppressor That Inhibits Papillary Thyroid Carcinoma Cell Migration and Invasion. *International Journal of Endocrinology*, 2020.

Yu, Y., Liu, H., Ikeda, Y., Amiot, B.P., Rinaldo, P., Duncan, S.A. and Nyberg, S.L., 2012. Hepatocyte-like cells differentiated from human induced pluripotent stem

cells: Relevance to cellular therapies. *Stem Cell Research*, 9(3), pp.196–207. NIH Public Access.

Yuan, F., Yu, Y., Zhou, Y.L. and Zhang, X.X., 2020. 5hmC-MIQuant: Ultrasensitive Quantitative Detection of 5-Hydroxymethylcytosine in Low-Input Cell-Free DNA Samples. *Analytical Chemistry*, 92(1), pp.1605–1610.

Yuan, S., Li, L., Xiang, S., Jia, H. and Luo, T., 2019. Cadherin-11 is inactivated due to promoter methylation and functions in colorectal cancer as a tumour suppressor. *Cancer Management and Research*, 11, pp.2517–2529.

Zarour, L.R., Anand, S., Billingsley, K.G., Bisson, W.H., Cercek, A., Clarke, M.F., Coussens, L.M., Gast, C.E., Geltzeiler, C.B., Hansen, L., Kelley, K.A., Lopez, C.D., Rana, S.R., Ruhl, R., Tsikitis, V.L., Vaccaro, G.M., Wong, M.H. and Mayo, S.C., 2017. Colorectal Cancer Liver Metastasis: Evolving Paradigms and Future Directions. *CMGH*, 3(2), pp.163–173. Elsevier Inc.

Zhang, J., Fan, J., Venneti, S., Cross, J.R., Takagi, T., Bhinder, B., Djaballah, H., Kanai, M., Cheng, E.H., Judkins, A.R., Pawel, B., Baggs, J., Cherry, S., Rabinowitz, J.D. and Thompson, C.B., 2014. Asparagine plays a critical role in regulating cellular adaptation to glutamine depletion. *Molecular Cell*, 56(2), pp.205–218.

Zhang, J. and Zheng, Y.G., 2016. SAM/SAH Analogs as Versatile Tools for SAM-Dependent Methyltransferases. *ACS Chemical Biology*, 11(3), pp.583–597. American Chemical Society.

Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.L., Luo, C., Jiang, H. and He, C., 2012. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nature Chemical Biology*, 8(4), pp.328–330.

Zhang, L. ying, Li, P. ling, Wang, T. zhen and Zhang, X. chen, 2015. Prognostic values of 5-hmC, 5-mC and TET2 in epithelial ovarian cancer. *Archives of Gynecology and Obstetrics*, 292(4), pp.891–897.

Zhang, L.Y., Han, C.S., Li, P.L. and Zhang, X.C., 2016. 5-Hydroxymethylcytosine expression is associated with poor survival in cervical squamous cell carcinoma. *Japanese Journal of Clinical Oncology*, 46(5), pp.427–434.

Zhang, P., Huang, B., Xu, X. and Sessa, W.C., 2013. Ten-eleven translocation (Tet) and

thymine DNA glycosylase (TDG), components of the demethylation pathway, are direct targets of miRNA-29a. *Biochemical and Biophysical Research Communications*, 437(3), pp.368–373.

Zhang, P., Zuo, Z., Wu, A., Shang, W., Bi, R., Jin, Q., Wu, J. and Jiang, L., 2017. miR-600 inhibits cell proliferation, migration and invasion by targeting p53 in mutant p53-expressing human colorectal cancer cell lines. *Oncology Letters*, 13(3), pp.1789–1796.

Zhang, T., Guan, X., Choi, U.L., Dong, Q., Lam, M.M.T., Zeng, J., Xiong, J., Wang, X., Poon, T.C.W., Zhang, H., Zhang, X., Wang, H., Xie, R., Zhu, B. and Li, G., 2019. Phosphorylation of TET2 by AMPK is indispensable in myogenic differentiation. *Epigenetics and Chromatin*, 12(1).

Zhang, W., Lu, Z., Gao, Y., Ye, L., Song, T. and Zhang, X., 2015. MiR-520b suppresses proliferation of hepatoma cells through targeting ten-eleven translocation 1 (TET1) mRNA. *Biochemical and Biophysical Research Communications*, 460(3), pp.793–798.

Zhang, W., Xia, W., Wang, Q., Towers, A.J., Chen, J., Gao, R., Zhang, Y., Yen, C. an, Lee, A.Y., Li, Y., Zhou, C., Liu, K., Zhang, J., Gu, T.P., Chen, X., Chang, Z., Leung, D., Gao, S., Jiang, Y. hui and Xie, W., 2016. Isoform Switch of TET1 Regulates DNA Demethylation and Mouse Development. *Molecular Cell*, 64(6), pp.1062–1073.

Zhang, Y.W., Wang, Z., Xie, W., Cai, Y., Xia, L., Easwaran, H., Luo, J., Yen, R.W.C., Li, Y. and Baylin, S.B., 2017. Acetylation Enhances TET2 Function in Protecting against Abnormal DNA Methylation during Oxidative Stress. *Molecular Cell*, 65(2), pp.323–335.

Zhao, B., Yang, Y., Wang, X., Chong, Z., Yin, R., Song, S.-H., Zhao, C., Li, C., Huang, H., Sun, B.-F., Wu, D., Jin, K.-X., Song, M., Zhu, B.-Z., Jiang, G., Rendtlew Danielsen, J.M., Xu, G.-L., Yang, Y.-G. and Wang, H., 2014. Redox-active quinones induces genome-wide DNA methylation changes by an iron-mediated and Tet-dependent mechanism. *Nucleic Acid Research*, 42(3).

Zhao, C., Setrerrahmane, S. and Xu, H., 2015. Enrichment and characterization of cancer stem cells from a human non-small cell lung cancer cell line. *Oncology*

Reports, 34(4), pp.2126–2132.

Zhao, W., Song, M., Zhang, J., Kuerban, M. and Wang, H., 2015. Combined identification of long non-coding RNA CCAT1 and HOTAIR in serum as an effective screening for colorectal carcinoma. *International Journal of Clinical and Experimental Pathology*, 8(11), pp.14131–14140.

Zhao, X., Liu, X., Wang, G., Wen, X., Zhang, X., Hoffman, A.R., Li, W., Hu, J.F. and Cui, J., 2016. Loss of insulin-like growth factor II imprinting is a hallmark associated with enhanced chemo/radiotherapy resistance in cancer stem cells. *Oncotarget*, 7(32), pp.51349–51364.

Zhavoronkov, A. and Bhullar, B., 2015. Classifying aging as a disease in the context of ICD-11. *Frontiers in Genetics*, 6(NOV), p.326.

Zhavoronkov, A. and Moskalev, A., 2016. Editorial: Should We Treat Aging as a Disease? Academic, Pharmaceutical, Healthcare Policy, and Pension Fund Perspectives. *Frontiers in Genetics*, 7(FEB), p.17.

Zhou, Y., Xia, L., Wang, H., Oyang, L., Su, M., Liu, Q., Lin, J., Tan, S., Tian, Y., Liao, Q. and Cao, D., 2018. Cancer stem cells in progression of colorectal cancer. *Oncotarget*, 9(70), pp.33403–33415.

Zhou, Z., Zhang, H.-S., Liu, Y., Zhang, Z.-G., Du, G.-Y., Li, H., Yu, X.-Y. and Huang, Y.-H., 2018. Loss of TET1 facilitates DLD1 colon cancer cell migration via H3K27me3-mediated down-regulation of E-cadherin. *Journal of Cellular Physiology*, 233(2), pp.1359–1369.

Zhu, F., Zhu, Q., Ye, D., Zhang, Q., Yang, Y., Guo, X., Liu, Z., Jiapaer, Z., Wan, X., Wang, G., Chen, W., Zhu, S., Jiang, C., Shi, W. and Kang, J., 2018. Sin3a-Tet1 interaction activates gene transcription and is required for embryonic stem cell pluripotency. *Nucleic Acids Research*, 46(12), pp.6026–6040.

Zhu, G., Li, Y., Zhu, F., Wang, T., Jin, W., Mu, W., Lin, W., Tan, W., Li, W., Street, R.C., Peng, S., Zhang, J., Feng, Y., Warren, S.T., Sun, Q., Jin, P. and Chen, D., 2014. Coordination of engineered factors with TET1/2 promotes early-stage epigenetic modification during somatic cell reprogramming. *Stem Cell Reports*, 2(3), pp.253–261.

Zhu, H., Wang, G. and Qian, J., 2016. Transcription factors as readers and effectors of DNA methylation. *Nature Reviews Genetics*, 17(9), pp.551–565.